

colc - H02801-7-0325081

AMERICAN PHILOSOPHICAL QUARTERLY



Edited by

NICHOLAS RESCHER

With the advice and assistance of the Board of Editorial Consultants:

Alan R. Anderson

Kurt Baier

Stephen F. Barker

Monroe Beardsley

Nuel D. Belnap, Jr.

Roderick M. Chisholm

L. Jonathan Cohen

James Collins

James M. Edie

José Ferrater-Mora

Richard M. Gale

Peter Thomas Geach

Adolf Grünbaum

Carl G. Hempel

John Hospers

Raymond Klibansky

Hugues Leblanc

Ernan McMullin

Benson Mates

John A. Passmore

Richard H. Popkin

Richard Rorty

George A. Schrader

Michael Scriven

Wilfrid Sellars

Alexander Sesonske

Manley H. Thompson, Jr.

John W. Yolton

Volume 7/Number 1

JANUARY 1970

CONTENTS

- | | | | |
|--|----|---|----|
| I. PETER K. MACHAMER: <i>Recent Work on Perception</i> | 1 | V. MARCUS B. HESTER: <i>Purpose in Painting and Action</i> | 62 |
| II. RICHARD B. BRANDT: <i>Traits of Character: A Conceptual Analysis</i> | 23 | VI. ROBERT BROWN: <i>The Burden of Proof</i> | 74 |
| III. MICHAEL R. AYERS: <i>Substance, Reality, and the Great, Dead Philosophers</i> | 38 | VII. STORRS MCCALL: <i>A Non-classical Theory of Truth, With an Application to Intuitionism</i> | 83 |
| IV. AMELIE RORTY: <i>Plato and Aristotle on Belief, Habit, and Akrasia</i> | 50 | <i>Books Received</i> | 89 |

AMERICAN PHILOSOPHICAL QUARTERLY

POLICY

The *American Philosophical Quarterly* welcomes contributions by philosophers of any country on any aspect of philosophy, substantive or historical. However, sufficient articles will be published, and not news items, book reviews, critical notices, or "discussions."

MANUSCRIPTS

Contributions may be as short as 2,000 words or as long as 25,000. All manuscripts should be typewritten with wide margins, and at least double spacing between the lines. Footnotes should be used sparingly and should be numbered consecutively. They should also be typed with wide margins and double spacing. The original copy, not a carbon, should be submitted; authors should always retain at least one copy of their articles.

COMMUNICATIONS

Articles for publication, and all other editorial communications and enquiries, should be addressed to: The Editor, *American Philosophical Quarterly*, Department of Philosophy, University of Pittsburgh, Pittsburgh, Pennsylvania 15213.

OFFPRINTS

Authors will receive gratis two copies of the issue containing their contribution. Offprints can be purchased through arrangements made when checking proof. They will be charged for as follows: The first 50 offprints of 4 pages (or fraction thereof) cost \$12, increasing by \$1 for each additional 4 pages. Additional groups of 50 offprints of 4 pages cost \$8, increasing by \$1 for each additional 4 pages. Covers will be provided for offprints at a cost of \$4 per group of 50.

SUBSCRIPTIONS

The price *per annum* is eight dollars for individual subscribers and fourteen dollars for institutions. Checks and money orders should be made payable to the *American Philosophical Quarterly*. All back issues are available, and are sold at the rate of three dollars to individuals, and four dollars to institutions. All correspondence regarding subscriptions and back orders should be addressed directly to the publisher (Basil Blackwell, Broad Street, Oxford, England).

MONOGRAPH SERIES

The *American Philosophical Quarterly* also publishes a supplementary Monograph Series. Apart from the publication of original scholarly monographs, this includes occasional collections of original articles on a common theme. The current monograph is included in the subscription, and past monographs can be purchased separately but are available to individual subscribers at a substantially reduced price. The back cover of the journal may be consulted for details.

335051



AMERICAN PHILOSOPHICAL QUARTERLY

Edited by
NICHOLAS RESCHER

With the advice and assistance of the Board of Editorial Consultants:

Virgil C. Aldrich
Alan R. Anderson
Kurt Baier
Stephen F. Barker
Monroe Beardsley
Nuel D. Belnap, Jr.
Roderick M. Chisholm
L. Jonathan Cohen
James Collins
Arthur C. Danto

James M. Edie
José Ferrater-Mora
Richard M. Gale
Peter Thomas Geach
Adolf Grünbaum
Carl G. Hempel
John Hospers
Raymond Klibansky
Hugues Leblanc
Ernan McMullin

Benson Mates
John A. Passmore
Richard H. Popkin
Richard Rorty
George A. Schrader
Michael Scriven
Wilfrid Sellars
Alexander Sesonske
Manley H. Thompson, Jr.
John W. Yolton



VOLUME 7 (1970)

PUBLISHED BY BASIL BLACKWELL WITH THE COOPERATION OF THE UNIVERSITY OF PITTSBURGH

AMERICAN PHILOSOPHICAL QUARTERLY

CONTENTS OF VOLUME 7 (1970)

AYERS, MICHAEL R.: <i>Substance, Reality, and the Great, Dead Philosophers</i>	38
BEATTY, JOSEPH: <i>Forgiveness</i>	246
BENNETT, JONATHAN: <i>The Difference Between Right and Left</i>	175
BEROFSKY, BERNARD: <i>Purposive Action</i>	311
BRANDT, RICHARD B.: <i>Traits of Character: A Conceptual Analysis</i>	23
BROWN, ROBERT: <i>The Burden of Proof</i>	74
CARR, DAVID: <i>Husserl's Problematic Concept of the Life-World</i>	331
COHEN, L. JONATHAN: <i>Some Applications of Inductive Logic to the Theory of Language</i>	299
<i>Corrigenda</i>	372
DORE, CLEMENT: <i>God, "Soul-Making" and Apparently Useless Suffering</i>	119
GALE, RICHARD: <i>Negative Statements</i>	206
GEWIRTH, ALAN: <i>Must One Play the Moral Language Game?</i>	107
GIBBS, BENJAMIN: <i>Real Possibility</i>	340
HENDERSON, T. Y.: <i>In Defense of Thrasymachus</i>	218
HESTER, MARCUS B.: <i>Purpose in Painting and Action</i>	62
KNOX, JOHN, JR.: <i>Does Becoming Entail a Contradiction?</i>	357
KYBURG, HENRY E.: <i>On a Certain Form of Philosophical Argument</i>	229
LESLIE, JOHN: <i>The Theory That the World Exists Because It Should</i>	286
MACHAMER, PETER K.: <i>Recent Work on Perception</i>	I
MARGOLIS, JOSEPH: <i>Egoism and the Confirmation of Metamoral Theories</i>	260
MARTIN, REX: <i>On the Logic of Justifying Legal Punishment</i>	253
MCCALL, STORRS: <i>A Nonclassical Theory of Truth, With an Application to Intuitionism</i>	83
MICHALOS, ALEX C.: <i>Rational Decision Making in Committees</i>	91
OPPENHEIM, FELIX: <i>Egalitarianism as a Descriptive Concept</i>	143
RORTY, AMELIE: <i>Plato and Aristotle on Belief, Habit, and Akrasia</i>	50
RORTY, RICHARD: <i>Wittgenstein, Privileged Access, and Incommunicability</i>	192
SANFORD, DAVID: <i>Disjunctive Predicates</i>	162
SHOEMAKER, SYDNEY: <i>Persons and Their Pasts</i>	269
SIMPSON, EVAN: <i>Actions and Extensions</i>	349
SKYRMS, BRIAN: <i>Return of the Liar: Three-Valued Logic and the Concept of Truth</i>	153
SPARSHOTT, FRANCIS E.: <i>Disputed Evaluations</i>	131
VISION, GERALD: <i>Essentialism and the Sense of Proper Names</i>	321
WALTER, EDWARD F.: <i>Empiricism & Ethical Reasoning</i>	364
WOLTERSTORFF, NICHOLAS: <i>Objections to Predicative Relations</i>	238

I. RECENT WORK ON PERCEPTION

PETER K. MACHAMER

THE paramount problem of perception is the relation between perception and knowledge. The solution of this problem necessitates first getting clear about the nature of perception, and finding in what way perception can best be analyzed. I propose to deal with various recent works, works which are selected for their quality and for the particular position maintained in them. The works considered fall into the following broad classifications: sense-data positions; sense-data criticisms; "harmless" sense-data theories; realistic theories; and psychological theories. I have tried, through the discussion of particular works, to show the problems that lie within a whole class.

Many works are included in the bibliography which are not mentioned in the essay. Some have not been discussed because they fall outside of the main theme I have chosen—for example, H. P. Grice's "Some Remarks About the Senses." There are others which duplicate or are very similar to works I have discussed. No doubt, there are some I have just overlooked. And there is a final group for which silence is the best possible comment.

I. SENSE-DATA I: PRO

Since the turn of the century much of the literature has centered around the concept of sense-data. There are, at this writing, only a few staunch supporters upholding the sense-data position in anything approaching its classic form. (The most cogent of these men, Casimir Lewy, has yet to publish anything on the subject; it is to be hoped that he will do so soon.) The sense-data position in its simplest form can be seen as an answer to a question about the nature of perception, i.e., what do we perceive?

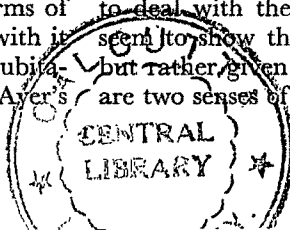
It would be best to sketch briefly the sense-data position in its traditional form. This position holds that whenever one perceives something, one is perceiving (or sensing) a sense-datum. The argument for this claim is usually one of the forms of the argument from illusion; and conjoined with it is usually a claim about the certainty or indubitability of the datum sensed. That is, in Ayer's

language, while we can make a mistake using our everyday language of physical objects we cannot make a mistake if we confine ourselves to talk of sense-data. When we perceive something, at the sense-datum level it must be as it appears to be. As it turns out this claim for the certainty of sense-datum statements is crucial. Not only does it function as an indubitable basis for knowledge, i.e., the sense-datum theorist's solution to the basic problem, but it gives such a theorist a way of accounting for illusions and hallucinations and the obstinate remarks and reports of those who have them.

Benson Mates in his article "Sense-Data" explicitly neglects the fundamental epistemological problem of the relation between perception and knowledge (pp. 228 and 236). He regards the introduction of "sense-data" as an attempt to give the denotation, not the sense, of the term. The point behind taking this Fregean tack is to escape the criticisms of unclarity brought by Quine, Paul, Ryle, and Austin by showing that the unclarity of the expression "sense-data" is somewhat irrelevant to understanding the claim that sense-data exist. Mates's most interesting section is about the introduction of terms in these two ways, i.e., by sense and by denotation.

Mates argues that there are two senses of "see;" one of which is a private sense-datum sense, while the other is public. His argument runs as follows: The sentence "Smith and Jones both see the Campanile" is both true and false. It is true *ex hypothesi* in the public sense; while it is false in the sense where what is seen marks off a private object of perception which varies with the perspective and bodily constitution of the perceiver. Since no sentence can be both true and false in the same sense, Mates concludes that there must be two senses of "see."

This argument leads to many problems which Mates does not recognize and he gives no hint how to deal with them. First, the argument does not seem to show that there are two senses of "see," but rather, given his Fregean semantics that there are two senses of the expression "the Campanile";



one where the referent is a private object of perception, and the other a public building. Also if we note the reason for thinking that there is a private sense of the sentence in question, i.e., the differing bodily constitutions and perspectives of the observer, we can note that the only way in which this reason can be established is by perception (of the requisite bodies, responses, behavior patterns, etc.). But from the observation that each of the objects denoted by the subject terms can be an object of perception it follows that the reference of "Smith" and "Jones" is liable to vary just as the reference of "the Campanile." Now given the three expressions each of which has two possible types of referents, we see that there are not two senses to the sentence but rather eight. The main problem then for Mates is to give some account of these multifarious senses. Which one is the sense-datum sense? Some account is needed to explain the relations between public objects and private objects. Such an account would have to answer the standard arguments about privacy. In sum, Mates has not yet succeeded in giving a coherent account of the possibility of sense-data.

A most interesting feature of Mates's article is his insistence that the usefulness of sense-data is irrelevant to the question of whether they exist or not. In effect his point is that even if there is no reason to use a sense-datum language it may still be true that sense-data exist, i.e., that there are such recognizable things as sense-data. Of course, on Mates's agnostic stand about the epistemological function of sense-data it is hard to understand how the issue of their existence would ever get raised, and why, even if he were right, anyone should bother to carry on with it.

Another proponent of the sense-data position is A. J. Ayer (*The Problem of Knowledge* and "Has Austin Refuted the Sense-Datum Position?"). Ayer has greatly qualified his earlier views (as found in *Foundations of Empirical Knowledge*) so that he almost fails to take any position at all. But he comes through still plumping for the two languages view; a language of material objects and a language of sense-data. In the revised Ayer this duality turns out to be the counterpart of the traditional theoretical-observational distinction. For Ayer, speaking of physical objects is a way of interpreting our sense experiences. In other words the physical object language is essentially a theory constructed to account for data of the senses, and being theoretical in character it transcends the senses themselves, i.e., it is not simply a logical

construction (*Problem*, p. 132 and "Has Austin, etc.," p. 119 ff.). A nice criticism of this whole approach, which can be found earlier in Ernst Mach, is available in Peter Alexander's *Sensationalism and Scientific Explanation*. Parts of the line he develops there will be mentioned below in various places.

Ayer still gives an argument from illusion, but not nearly as neatly as earlier. The newer one runs:

1. Because of the possibility of illusion, it is not necessarily true that whenever I seem to be perceiving something I really am perceiving it.

2. Minimally, it follows from an instance of (1) that if I seem to see a cigarette case, then it seems to me that I see a cigarette case.

3. I am now seeing a seeming cigarette case.

4. Applying steps (1)-(3) to all cases of perception we get, whenever anyone perceives or thinks he perceives a physical object, he must be perceiving a seeming object (pp. 96-97).

Ayer himself argues against part of the argument above, drawing attention especially to the weakness of the "usual" way of moving from (2)-(3). But in the *Problem of Knowledge* he seems to maintain that there is at least one sense in which the argument goes through, i.e., a sense in which the existence of an object is not entailed by a claim to see it and in which it is not possible to doubt what I am seeming to see. (In the later paper he backs off of the indubitability claim; cf. pp. 137-138.) In other words he holds that there is a sense of "seeming" in which whenever one perceives or thinks he perceives there is something which seems to him to be perceived (*Problem*, pp. 100-104). Of course, what it is, this something that is perceived in all cases is a sense-datum.

Apart from the problems of the indubitability claim (which will be discussed below), Ayer inherits all the tiresome difficulties which surround the theory-observation distinction in a positivistically oriented philosophy of science. The many difficulties surrounding the attempts to mark out the theory-observation distinction have been raised in many places. N. R. Hanson, in *Patterns of Discovery*, brought out the way in which many of the interesting cases in science depend upon observations which are heavily theory-laden. This idea has also been taken up and refined by Peter Alexander (cf. above) in *Sensationalism and Scientific Explanation*; he there attacks a reductivist

program by showing the impossibility of a "pure" sense language. Finally, Peter Achinstein in "The Problem of Theoretical Terms" shows the necessity of relativizing the distinction to such an extent that one wonders of what use it could be. There have been recent attempts by Dudley Shapere and others at Chicago to construct a theory of science which does away with the traditional distinction completely. All this, of course, bodes badly for Ayer.

II. SENSE-DATA II: CONTRA

The sense-data position that I have been speaking about is often called "phenomenalism"; though it is of note that some philosophers, e.g., Ayer, feel they can uphold a sense-data theory without holding a phenomenalist position. The reason for this is that the traditional position of phenomenalism held that all statements about physical objects could in principle be reduced to statements about sense-data (or ideas, etc.). By 1955 this reducibility claim has almost universally been given up. So, for example, when Frank Ebersole goes to attack the sense-datum theorists ("On Seeing Things") he feels he only needs to argue against the much more restricted claim that sense-datum statements (or statements of immediate experience) are part of the evidence for claims about seeing physical objects.

Besides the condition mentioned above Ebersole sets out four other conditions which he claims a sense-data position wants to hold: such statements must be about present-experience, i.e., they must be "insulated in time"; they must not imply anything about the material world, i.e., they must be "insulated from the material world"; they must be private (which according to Ebersole does not mean that they must be absolutely incorrigible); and they must be implied by statements such as "I see x ." Ebersole goes on to make out a case that the only statements that even seem to fit all five conditions are those reporting hallucinations. But in the end, it turns out on Ebersole's view that hallucinatory statements (i.e., statements about hallucinations) are not really to be taken as perceptual statements but rather as reports of the experience the hallucinator is undergoing. That is, when someone says "I see red rats" what is being accomplished is that the hearer is being informed that the speaker is having an hallucination. The only analysis, he claims, of such hallucinatory statements is in terms of the whole context of

utterance; and when such an analysis is given the statement, unlike normal perception statements, implies the nonexistence of the object mentioned, e.g., if you know that I am hallucinating you are entitled to conclude that the red rats of which I speak do not exist (or minimally, that I have no evidence for saying they do). From this Ebersole draws the conclusion that the five conditions he has outlined are never fulfilled by statements of immediate experience or of any other sort. In fact he says "seeing" in the way one does when hallucinating is different from seeing things.

The problem here, of course, is that it is not at all clear why hallucinations need to be analyzed in the way Ebersole claims. It would seem that to do justice to the experience of hallucinating one should be able to give some account of the seeing that goes on in such an experience. Of course it's different from normal seeing, that's why its hallucinating. But to remark this difference is not to give an analysis of it, nor to go very far toward explaining why one is tempted to describe it as a case of seeing something. It is also in some ways like seeing objects, which is why people sometimes are confused and make mistakes in such cases.

Austin, in *Sense and Sensibilia*, is busy attacking three works published before our period began. He takes on the argument of illusion at first and argues that the conclusion that all one sees are sense-data trades on a confusion between a delusion in which something is conjured up, something unreal and private, and an illusion in which there is something really present and public. By conflating these two ideas one gets as a conclusion that what is really perceived is something which is really there and immaterial. Also Austin points out that sense-data are introduced to have a name for what we see. But if we had not lumped many different kinds of cases all together we would not be tempted to seek a single name for them all, i.e., "sense-data," but rather would be quite content with the names already in parlance, i.e., "hallucination," "mirage," etc. Since these names are more precise and informative, why do we need to seek another?

The important idea of "trouser-words" I shall leave almost undiscussed here. In general Austin's move is to point out to the sense-datum claimant that things are not as simple as he would like, e.g., that there are many ways that "looks like" is related to "is," and that there are many different uses of "see," not just two.

Of major importance in Austin's book is his

argument against the claim of certainty, or the "pursuit of the incorrigible." To beat this bugbear he takes off against Ayer's idea of the two languages. The claim is that there are not two languages, but only one; and that one is the one which is learned in childhood by all normal speakers of a language. Paraphrasing and elaborating this point about language which is found in Austin (and in various aspects in the article by F. H. George, Austin's "Other Minds," work by Bruner, Israel Scheffler's *Science and Subjectivity*, and the Strawson and Hampshire symposium "Perception and Identification") the argument against certainty might be sketched as follows: Whenever one applies a predicate to a particular, or more generally describes any experience, one is relying upon memory and identification. This fact about using predicates is partly the result of how we learn our language, and partly the result of what a predicate is like. We learn to use predicates (or concepts) by applying them to different particulars, and it's not hard to see that memory plays an important part in this, e.g., I must remember that it was the word "red" which I applied to this apple the first time, and perhaps that it was the color I was concerned with, in order to know that I should here and now apply the word "red" to what I see. Of course, the part played by memory need not be conscious (cf. parts below on Hirst and Bruner). Also since the very concept of a predicate involves knowing that certain particulars and not others fall under it, there is memory involved in the identification (or categorizing) of the particulars to which we want now to apply the predicate. (Cf. Strawson, Shoemaker, and Coburn).

As Austin says, whenever memory is involved we know that it is possible to make a mistake ("Other Minds," section on Sureness and Certainty). That is, to the object before me I might apply the word "red," i.e., I might say "This is red." But I might be mistaken in such an application. Likewise, I might make the less substantial claim "It now looks to me to be heliotrope;" here too one might show me my mistake by pointing out a patch of heliotrope. These mistakes seem on a par, if the latter is merely "verbal" as Ayer might claim then the former is so also, and indeed so are most of our mistakes.

The only way out of this criticism seems for the sense-datum theorist to insist upon a private language claim, where "heliotrope" takes on a private and ineffable meaning for me alone. But

how could we ever teach such a language? Is such a language even possible? If we answer it is not possible, the problem is solved for such a second language cannot exist; if one disagrees then some account is needed as to how others are somehow enabled to understand the private utterances of one who says "I see red rats playing Mozart." But this is not the place to get into the problems of privacy.

The main contention of Austin and those who have followed him is that sense-data are not needed (a claim which Mates argued to be irrelevant). We can speak perfectly adequately and intelligibly, saying everything we need, without having recourse to such peculiar entities as sense-data. Put positively the claim is that we ordinarily speak about seeing physical objects except for special purposes or when we are mistaken or fooled. This claim is taken over by the realists (or perhaps, more accurately, shared by them) and we shall discuss it below.

Once the certainty claim has been attacked and successfully felled, it is still open to a sense-datum theorist to attempt resuscitation by introducing a weaker form of sense-data. But it is important to note that while this move might be made (and has been by some), historically, at least, the temptation for introducing sense-data has been the certainty claim, i.e., the providing of an indubitable foundation for knowledge. Once such a claim is given up a large part of the appeal of such a position goes with it. Also when the certainty claim is given up, the sense-data position moves into what I shall call the "harmless view of sense-data" in order to get across the idea of its decreased ontology.

III. SENSE-DATA III: INTENTIONALITY AND THE "HARMLESS" VIEW

N. R. Hanson (*Patterns of Discovery*) put one aspect of this position in a way when he claimed that while there were no *purely* phenomenal statements there certainly were more or less phenomenal statements. And this is one line of thought which has served to introduce the idea of harmless sense-data. The other tack which needs to be considered is taken by those who propose to treat the objects of perception in a grammatical way, or, as Kenny and Chisholm have put it, to treat perception as having an intentional or formal object.

Chisholm has attempted to set out the conditions

for intentional objects in his book, *Perceiving* (Chap. II); he tried there to uphold Brentano's thesis that all psychological verbs are characterized by having intentional objects. Chisholm claims that probably the verb "perceive" takes an intentional object. The end purpose of Chisholm's claim here is that intentional verbs cannot be analyzed into non-intentional ones, and, thus, that perception, in its intentional character, is somehow primitive. To make clear what kind of an object it is Chisholm suggests three conditions which such objects must satisfy (he has since changed his mind, cf. his article "Intentionality" in *Encyclopedia of Philosophy*): Such objects may or may not exist (the non-existence condition); the sentences in which such objects occur need not imply either the truth or falsity of the propositional clause in that sentence (the indeterminacy condition); and finally, such sentences are referentially opaque, i.e., they do not preserve truth under all substitutions of equivalent names or descriptions for the object in question (the non-substitutivity condition). It is interesting to remark that at least two of the above conditions (the first two) have often been put forward explicitly as characteristics of sense-data.

G. E. M. Anscombe has put forward an account of sensation and perception which uses the notion of an intentional object. Anscombe sets out a non-existence condition, an indeterminacy condition, and a non-substitutivity condition which are almost the same as Chisholm's. She goes on to argue that the sense-datum theorists have taken the object in perception—i.e., what would be used to answer the question: what do we perceive?—materially in all cases and have thus committed a fallacy (cf. Chisholm, *Perceiving*, pp. 151-153). The only way, she claims, that perception is always of something is that there must always be an intentional object—there need not always be a material object, e.g., in hallucinations. But to say that seeing always has an intentional object is just to remark a feature of the grammar of perception and sensation statements for Anscombe. The problems that remain are especially the questions involving the relations between material objects and intentional objects in "veridical" seeing. Are there two objects? How exactly are they related? More generally it is unclear what can be drawn from the thesis that sensation statements always have grammatical objects. In the older terminology one is unclear as to what their ontological status is? And if they have none, how is the

pointing out of their character meant to clarify the problems about hallucinations and other nonmaterial cases of seeing which sense-data were introduced to solve?

I think that from the above survey one can grasp the idea that in the area of perception these intentional objects are, in fact, co-extensive with the sense-data often argued for. But since the intentional objects are not given, at least on Anscombe's account, any epistemological priority (but cf. Chisholm, pp. 184-185), they constitute one way of reworking the idea of a sense-datum which forgoes the bugbears so heavily laid upon by Austin and others. Unfortunately, it is hard to see why they are philosophically interesting with respect to perception and what exactly is meant to be their role in understanding the concept of perception—unless it is, as Chisholm hints, a way of showing that in one essential respect they are unanalyzable.

There is another tack to the harmless view of sense-data which can be found exemplified in the works of Hampshire, Ebersole, and Kneale. This view has some adherents in the camp of the causal theorists (e.g., Grice and Hirst) but I shall postpone discussion of these until the next section.

Anthony Quinton, in "The Problem of Perception" (pp. 66-68 in the Warnock volume) has allowed the possibility of a use of "appears" which is not particularly a guarded way of speaking (which doesn't suggest doubt), and is neither incorrigible or somehow fundamental to an account of knowledge. He also maintains that there is little use for such an expression.

It is to be noticed that the nature of sense-data here is quite different from what is traditionally claimed. This point is clearly brought out by William Kneale in "What Can We See?" where he argues for the use of the concept of a "view" as analytically helpful for perception. In Kneale's idea a view is something public, it corresponds to a visual field and is present every time one sees something. It is meant to be a noncommittal visual report (about the existence of the object). It is hard to imagine though on Kneale's account what "view" one has during an hallucination, and this seems a critical flaw.

I have mentioned only a few of the noncommittal or harmless notions of sense-data. What they all seem to have in common is that their sense-data are not certain and they are not in any sense more fundamental than physical object seeing. In fact, they are usually taken as being derivative from

the physical object use of perception verbs and are meant to be used in cases where caution or a particularly specific description is called for. It is somewhat common in psychology to find a like distinction drawn between perception where there must be an object present (an outside stimulus) and the reporting of appearances as in illusions and hallucinations, e.g., C. Solley and G. Murphy, *Development of the Perceptual World*, p. 17. The use of the noncommittal perceptual language often allows for a more guarded, more certain claim. Finally, it may also be used as Austin suggested in cases of doubt as to what is seen.

IV. THE CAUSAL THEORY AND REPRESENTATIONALISM

So far in this essay we have confined ourselves to speaking of sense-data which have been generated from the argument from illusion or its close kin. There is another argument for obtaining sense-data; and the arrived at entities may be either harmless or not depending upon the theorist. The simple theory of causality in perception, based upon neurological study, can be found in J. R. Smythies' *Analysis of Perception*. Smythies argues for what he and Lord Brain have called the television theory of perception. Roughly, the theory holds that sense-data are to physical objects as what is seen on a television screen is to the actors before the camera in the studio. That is, the neurological mechanisms which are used in perception and the light rays reflected from the object are mediate between the object itself and the image one sees (just as the camera and the radio waves are mediate between the actors and the home screen image). What we directly see are sense-data which are causally linked to physical objects. Smythies argues that they represent physical objects, and that by performing suitable mapping operations we can assure ourselves that the physical objects are like the sense-data we see. Others (Armstrong) have argued that measurement being more objective allows us through a deviate derivation to conclude that physical objects are truly represented by what we see, i.e., that there is a measurable correspondence between the sense-datum and the object. The problem with this position is that which Locke commentators have dwelt upon for years, for unless one begs the question there is no check independent of perception which allows us to examine the physical

objects to see whether they are represented accurately by what we see or not. There may be a possible out for this view on the grounds that the way physical objects really are can be independently checked by the physical parameters of particular equations, but no one has given such an argument and it would seem that the conclusion from such a move would be against a claim of representation. Also, the analogy to television is misleading and breaks down at the crucial point for whereas one can go to the studio and see the actors thereby bypassing the mediate steps, such a trip is not possible in the case of perception. In sum, this type of theory makes physical objects essentially unknowable, and as Hirst has pointed out the best that can be done for them is to make the sense-data into real entities which are just a duplication of what we have already in the physical objects; a move which is ontologically repulsive (R. J. Hirst, *The Human Senses and Perception*).

Another type of causal theory has been put forward by H. P. Grice ("The Causal Theory of Perception"). Grice takes seriously the claim of Austin and his followers that the word "appears" when used in normal contexts implies a doubt or denial of what is claimed to be seen. He then goes on to elaborate a theory of implication whereby it becomes possible in some cases either to detach or cancel what is normally implied. The outcome being that one might make a harmless sense-datum statement (without doubt or denial) as follows: "It looks to me red but I don't doubt or deny that it is red." Having thus vindicated the appearance statement from its Austinian objectors Grice goes on to sketch his causal theory. For a statement of appearance to be a necessary and sufficient condition for someone's perceiving an object the appearance must be causally dependent upon some sort of affair involving that object. But, Grice claims, the philosopher need not undertake the job of specifying the manner of causal connection; this is a scientist's job. The philosopher may indicate the type of connection he has in mind by examples. That is, for an object to be perceived by someone it is necessary and sufficient that it should be causally involved in the production of the sense-data (sense-impression) *in the kind of way* in which, e.g., when I look at my hand (in good light, etc.) it is responsible for its looking to me as if there were a hand before me. Claims for perceiving physical objects are justified, if they need to be, by showing that the object in question needs to exist if the

sense-datum statement in question is to be causally accounted for.

Grice has made an attempt in his reworking of the causal theory to account for the relations between sense-data and physical objects. The reason it is classified as one of the harmless sense-data theories is that it is possible to hold a theory like Grice's without making any of the infallibility or epistemic priority claims that accompanied the traditional view. Grice though gives no indication how he proposes to account for the nature of physical objects, but presumably on his view this would be a scientist's problem. At any rate he does not fall into the trap mentioned earlier because he completely eschews any claims for representation. There are other adherents to the causal theory, but I propose only to discuss one more, R. J. Hirst; and him later in this paper.

V. REALISM

The standard alternative to the sense-data and representationalist theories is the realist position. Fundamentally this position claims that we directly are aware of physical objects in many cases of perception. We have already mentioned Austin, who along with Wittgenstein, may be taken as a proponent of one way of supporting Realism. Austin was in no doubt that most of the time we perceive physical objects, though he was of course much shrewder in this claim than that time-honored scapegoat, the naive-realist. The force of Austin's position comes out in two rather indirect ways; indirect, as he never directly argues for such a theory. First, the realist position is inherent in his defense of the common man, e.g., where he takes Ayer to task for implying that the common man is in need of a drastic revision of his beliefs. The implication is that the common man believes we usually see physical objects. Second, the argument for perception of objects can be found in his discussion of "real." In this he holds that the negative of "real" is the "trouser-word;" so that to say that we see appearances and not reality needs to be spelled out by making clear what the appearances are. Austin's argument is that in the sense-data position these words have lost their normal uses; since we only see appearances and logically never could see reality he holds that what it is to see an appearance makes no sense. It has been deprived of any use. In support he brings up more normal cases where "real" has a use, e.g., in "real cream," "a real

duck," etc. Here "real" functions usefully in marking off what is spoken about from the artificial or a decoy. The conclusion sought is that the sense-data claim against seeing physical objects makes little sense, and that we have already in our language all the tools we need (or at least a sufficient supply) to account for perception. I hope it is obvious that in this discussion Austin alternatively presupposes and indirectly argues for directly perceiving physical objects in standard types of cases (cf. also *Sense and Sensibilia*, pp. 14-20).

Another manner of setting up the realist position can be found in the writings of Smart, Armstrong, and Ellis. David Armstrong, in *Perception and the Physical World*, sets out a realistic view of perception in which perception is a direct confrontation between the perceiver and the world. Perceiving is the immediate acquiring of beliefs about the world as they work in a sense-data position. Perception provides us with knowledge of the physical world, as it were, without proof (p. 135). Armstrong's theory is an attempt to break down the distinction, which has been held off and on since Plato, between perceiving and judging.

Armstrong, since he takes science as showing us what is real, is forced to the interesting conclusion that there exists widespread illusion in most of our ordinary perceptions; since what the scientist sees and measures is not what we ordinarily talk about. He tries to reconcile this idea with ordinary language by demarcating two senses of "real:" the ordinary everyday use and the more proper and correct use of the scientist. He holds that these are connected usages, but never really makes out the case as to how they are connected. If one left aside talk of "real" there might be something to this view, but I shall take this up again in the final section of this essay.

Armstrong has been criticized by John Nelson ("D. M. Armstrong's Theory of Perception"). Nelson argues against Armstrong's equating perceiving and believing. For example, if perceiving is reducible to believing, and believing is propositional in character, then perceiving must be propositional in character. That is, if the object of a belief is a proposition then the object of a perception must also be a proposition. But the only things propositional that are related to the world are facts, so it must be that we perceive facts—not physical objects. This makes the concept of a physical object or material thing horribly mysterious. More generally, Hirst has

pointed out that all realist positions have problems in trying to account for the facts of hallucination and beliefs about them; especially the qualitative similarity so stressed by the sense-data people which would lead one to think that hallucinating is perceiving "veridically".

J. J. C. Smart (in *Philosophy and Scientific Realism*) argues against phenomenalism on the grounds that it is seeking some sort of incorrigible assertions upon which to base our knowledge. By destroying the idea of certainty he feels he has destroyed phenomenalism. Unfortunately, Smart doesn't give enough of an argument for his claims. Realism comes in for Smart in two ways: first, it seems he thinks that it is the only alternative to phenomenalism; second, he claims that realism can give a better account of scientific theories.

Finally, I take another representative of the realist group, Brian Ellis ("Physical Monism"). Ellis calls his position physical monism to separate it from the fundamentally dualistic positions of the sense-datum theorists and representationalists. He brings to bear his training in science in developing a theory along the lines mentioned by Armstrong. After a preamble of some rather loose talk about how to judge theories by their coherence with other theories, he plumps for physicalism as the theory of mind best suited to other theories for which we have independent evidence; presumably these are physiological theories and the like. He then discusses the process of perception and for a variety of not too clear reasons he accepts the idea that perception is just one of the ways of gaining knowledge. He rejects, in doing so, any theories that maintain a distinction between perception and inference; strangely he includes Armstrong in those rejected by this move.

Ellis sets out the physical monist's position as the belief that what is immediately perceivable is dependent upon the immediate neural responses we have acquired the capacity to make (p. 159). From this it follows for Ellis that we are immediately aware of spatial and temporal relationships, causal relationships, and relations of similarity, identity, and difference, and temporal sequence. In perception we can and do acquire concepts as we do in inference. Ellis holds that beliefs acquired by perception are only part of our beliefs and are neither pure nor basic. They are in fact influenced by the rational activities of which we are capable and the prior history of the perceiver's sensory stimulation.

In both Ellis and Armstrong we find a tendency

to take "stimulation" as a clear concept needing no explication. Certainly it can be marked out physically, but unless, as Ellis does, one holds an identity theory of mind such a demarcation would lead directly to the problems of representationalism. All of which is meant to point out that some realistic positions will stand or fall with their larger thoughts on the nature of mind: not a particularly surprising conclusion.

There is one strange fact about Ellis' paper which I have not yet mentioned. His scientific caution forces him to leave open the question of whether or not infants are born with an innate set of neural responses. But in discussing the genetic development he holds that his position applies only to the "trained" (i.e., normal adult) perceiver. He wants to hold that before a child can have any belief he must have a whole set of beliefs (pp. 159-160). In other words, there is no way that one could come to believe anything on this theory for it requires beliefs existent in order to form new beliefs. In order to make any sense of this it would seem that he must allow for innateness and some sort of maturational devices. More than this, the background set of beliefs must somehow be formed in some belief-like nebulous way. One is very unclear exactly what one is to say about children's perceptions on this view. Newborn infants perceive and discriminate at birth (cf. T. Bower, "The Visual World of Infants") and older children have set up a system of concepts which hold together in a manner far different from ours; a system in which perception plays a much more important part than later (cf. J. Piaget, *The Child's Conception of Number*). These may seem like side issues, but Ellis himself claims that a system of perception in so far as it is not empirical must stand or fall with how it relates to other theories for which we have evidence.

VI. R. J. HIRST'S THEORY OF PERCEPTION

There is one major theory of perception which I have mentioned at points throughout the above sections, but have yet to deal with in any specificity by itself. Hirst previewed his theory in *Problems of Perception*, but really developed it in a coherent way in a 95 page article appearing in *Human Senses and Perception*. This position can also be found in condensed form in his articles in the *Encyclopedia of Philosophy*. Hirst takes up what he calls a variant of the representative theory of perception. The main reason he has for holding this is that it

seems to him the only theory that accords with the neurological facts and the causal relationship between seeing and the world, and which can also adequately account for the facts of hallucinations and after-images. He tries, however, to get rid of the unnecessary and ontologically ugly duplication of the percept and the object. He manages this by a double-aspect theory of mind: where the mental image (what appears in perceptual consciousness) and the physical stimuli (which can be publicly observed) are but two aspects of the same phenomenon. They differ only by their mode of access, i.e., introspection and observation. Perceptual consciousness is that which we note before our eyes by introspection. It is immediate and undoubting at the time it is perceived, but may in fact prove erroneous. Perceptual consciousness also involves prior knowledge, beliefs, and attitudes which shape it and cause it to have the character that it does. Through this device, Hirst is able to accommodate in a positive manner psychological facts, e.g., object constancy, color variations, and the role of figure-ground in perception. Since conscious inference does not play a part in one's normal apprehension of what is seen, it follows, on Hirst's view, that the background of knowledge and belief must work unconsciously (though when we describe how it works we must do so as if it were conscious). In other words, Hirst gives what he calls a genetic explanation of perception in terms of unconscious activities as opposed to some sort of reductive analysis.

In spelling out this idea he makes use of the data available from the psychologists. It seems, though, that he makes at least one mistake. He distinguishes in the sensory process between the sensory activities and the modificatory activity; the sensory being related in some complex fashion to the stimuli pattern. The recent work of Lettvin and McCulloch (e.g., "What the Frog's Eye Tells the Frog's Brain") would seem to cast doubt upon this distinction. It seems that selection and categorizing are already present in the reception of the stimuli. Thus, Hirst must do some reworking on his distinction between the *percipienda* (stimulus properties) and what appears in perceptual consciousness. Also it will muddle up his nicely laid out scale of perceptual consciousness.

One nice thing about Hirst's theory is that it seems to have a completeness and coherence which other writers lack. He takes pains to account for all the puzzling phenomena which have bothered philosophers, and does a good job of making them

fit together. Also he has taken into account the problem of how prior knowledge, needs, etc., influence perception. This is a facet not often remarked, though Ellis also gave it some prominence. There are other problems with Hirst's theory. For example, Hirst is not explicit enough in laying out the exact character and pattern of these unconscious activities. This is probably due to his overlooking the cognitive psychologists. I shall attempt in the next section to sketch the theory of Jerome Bruner which would seem to serve to fill the gap in Hirst's work. More central philosophically is the question of the theory's dependence upon a theory of mind which is by no means uncontroversial.

VII. PSYCHOLOGY AND THE PHILOSOPHY OF PERCEPTION

There exists in philosophical circles a prominent group of philosophers who are convinced that philosophic problems can be divorced from scientific problems. They would seem to believe that every discipline has its own demarcated area and that those persons trained in one area should not stray into another. This view has proved seminal for at least two lines of thought about the problems of perception: the first is that there is no philosophical problem about perception. On this view when one has gotten clear of all the muddles philosophers have made, e.g., the sense-datum puzzles, one will come to realize that there exists no "logical" problem. The questions about perception are causal questions, and the answers are to be found by psychologists. What is called for is an explanation of how our beliefs are related to perception. Of course, philosophy has the peripheral task of making this position clear and of freeing philosophers from traditional dogmas. A position along this line is suggested by Anthony Quinton in his paper "The Problem of Perception."

The fact that the history of philosophy wholly fails to support the claim of this group would not constitute a refutation for them. It would rather be used to support their position, showing how ancient the dogmas were, and indicating the need for change. The idea is that philosophers are finally beginning to get clear about what they should have been doing all along. On this view a person could accept the bulk of Hirst's argument but then go on to claim that it was irrelevant to philosophy.

The second line of thought I mentioned from

this position is that traditionally espoused by H. H. Price and G. E. Moore. On this view there is a philosophical problem about perception, but it is different from any psychological problems. The adherents of this view claim that no psychological evidence or data is necessary, or even relevant, to the treatment of philosophical issues in perception. An interesting fence-straddling version of this position can be found in N. R. Hanson's *Patterns of Discovery* where he claims that since philosophy has only to do with conceptual matters it has nothing to do with fact, and then he goes on to suggest that the reading of some 20 psychological articles would much improve contemporary discussions of perception—but he never makes clear how (p. 181, footnote 1 for page 15). I am at a loss to see how one is sharply to separate facts from concepts; and indeed how one is to grasp concepts except by grasping facts. But I must beg off further discussion. An interesting discussion of some philosophers' attempts to dismiss psychology as relevant to perception can be found in J. A. Fodor's "Could There be a Theory of Perception?" Fodor comes to no staggering conclusions, except that philosophers had best slow down and consider what the psychologist is really about rather than, e.g., trying to dismiss learning theory with one quick wave of the arm of clarity.

I hope to sketch out briefly how parts of the "psychological" theory of Jerome Bruner might be used to augment and support Hirst's theory. In setting out his theory of perception Bruner ("On Perceptual Readiness") stresses an input-output model, and a process of unconscious categorical inference by which the input is converted into the final end of experience-as-we-know-it. Perception, for Bruner, is characterized by its categorical nature, i.e., how the different concepts and drives of a person play a part in perception; and by its inferential nature, i.e., how we learn the relations obtaining between objects and events in the world, how we learn category systems appropriate for such events, and how we learn to predict and check these former inferences. Perception involves the learning of appropriate categories, the learning of cues for placing objects in a system of categories, and learning various expectations about what objects are likely to occur in a given environment (p. 229—page numbers refer to the anthologized version; see bibliography).

He goes on to develop this theory stressing the aspects of cue utilization and category accessibility.

The former is relatively unproblematic, but the latter is most interesting. The idea of category accessibility is that different categories vary in terms of their accessibility, i.e., they vary with regard to the readiness with which the organism uses them to categorize given stimulus inputs. Such accessibility is said to depend upon two factors: the expectancies of the person with regard to a specific event's occurrence in a given situation, i.e., what one expects in this type of situation; and the search requirements of the organism imposed by his needs and ongoing enterprises, i.e., roughly, what he's looking for. The purpose of such accessibility rankings is to minimize the surprise value of the environment by keeping things relatively constant, and to maximize the attainment of sought-after objects and events (p. 235). The interest of such a theory of accessibility is that it helps to account for the traditional problems of illusion, and other "non-veridical" perceptions; and it does so in a manner which integrates such facts with a larger theory which itself correlates with wider ideas of learning theory and the like.

Finally, I should like to mention that in this paper Bruner also sketches an hypothesis about the mechanisms which might be thought to underlie such a process. A criticism which begins at this molecular level is found in the work of Lettvin and McCulloch. Lettvin has informally said that his work undercuts any simple model of the input-output type. The coding devices of sensory reception are so complex that there is no simple manner in which it can be schematically represented. (Cf. earlier reference to Lettvin and McCulloch). Now it may be that this work has ramifications only at the molecular level, and will leave untouched the overall scheme of Bruner. But if this work is applied accurately to man, complications would seem to be quite relevant to such a theory.

Before leaving this section, I want to draw attention to some good books which are relevant to this topic. A book, not written by a psychologist, rather by an art historian, but which contains a wealth of material and illustrations concerning the interrelations between perception and expectation and knowledge is E. H. Gombrich's *Art and Illusion*. It is a book that should be required of all students of perception. Also I want to call attention to a book written in a style which varies from the erudite to the introductory, but which contains many fascinating examples plus an interesting theory of the nature of illusions, i.e., R. Gregory,

Eye and Brain. Finally, an easily written review of much of the major literature in the psychology of perception can be found in M. D. Vernon's *The Psychology of Perception*.

VIII. PERCEPTION, KNOWLEDGE, AND SCIENCE

In this final section I want to treat a few works whose major themes are not about perception, but which, nonetheless, have much to say on the subject or which depend heavily upon some theory of perception. First the concept of perception and categorization has been used recently by Israel Scheffler (*Science and Subjectivity*) to try to mediate between the two major factions in the philosophy of science. Scheffler attempts to reconcile the older logical empiricist view of philosophy of science with the newer critics, e.g., Hanson, Feyerabend, and Kuhn. In the section on observation he criticizes a view of perception which relies upon an indubitable given (the view is that of C. I. Lewis), and also the idea that perception is theory-laden, which for Scheffler is tantamount to admitting it is entirely subjective. In arbitrating these two views his stated goal is to reintroduce the idea of objectivity into the philosophy of science. Throughout the work Scheffler does admirably in recounting positions and presenting both sides of the situation. He fails only at the crucial juncture of offering a satisfactory and coherent solution, e.g., it is patently obvious that Frege's distinction between "sense" and "reference" will not without much reworking serve as a basis for a judgement. Also it is hard to see how any simple pigeon-hole theory of mind can really reintroduce the "objectivity" so longingly sought.

But I believe that Scheffler is on to a good idea, and the way of approaching the problem is basically sound. It is desirable first to get clear about Hanson's position which Scheffler attacks; Hanson never claims that all seeing is theory-laden (though this is a very popular misconception). In fact he explicitly states at various points that unless there is some sense in which two people see the same thing there can be no interesting sense in which they see different things. Hanson, like Scheffler, is claiming that there is no *pure* phenomenal given, but he does admit the possibility (of which Scheffler seems unaware) that there are some statements which are more phenomenal than others.

In terms of theory of perception set out by

Hirst and supplemented by Bruner I believe it is possible to account for the differences in observing which Hanson wished to remark with his idea of theory-ladenness, and also to account for the objectivity which Scheffler seeks (though without having recourse to a rather unclear idea of natural categorization). More specifically, I should like to suggest that the idea of objectivity be sought in the idea of shared categories. Two men might disagree over a position in physics, but quite probably would share a system of color concepts and the like. The idea of sharing would have to be worked out along the genetic lines suggested by Hirst (and indeed earlier by Wittgenstein). That is, two persons would share a category when they learned the concept in the same way, i.e., have the same criteria for its application. Bruner's idea of accessibility is relevant also for it would be in terms of the idea of accessibility that we might be able to make sense out of conceptual changes in science, and also, more fundamentally, the idea of objectivity itself. The objective will be that which is agreed upon by those concerned in the field; note that "agreed" does not necessarily imply conscious agreement but probably more often an unconscious agreement of a criteriological sort. The exact level of objectivity can be further specified as the most accessible level of categories common to a particular group.

There are no doubt many problems involved in working out such an idea in detail, but such a project does seem to have some merit. It could be used to give a more specific content to many ideas currently floating about in the philosophic literature. For example, the idea of continuity of observation which can be found in Grover Maxwell's "The Ontological Status of Theoretical Terms" and Rom Harré's *Theories and Things* will fit nicely into this method of treatment. One, in effect, describes in detail the continuous changes of the concept of observation itself, remarking the interesting but mostly unmentioned shift that often occurs in science where the observations of a particular phenomenon cease to be problematic in themselves—i.e., where they are used as evidence or are subject to interpretive problems—and turn instead into data which are usable by all those who have learned the new concept.

The relevance of the problems of identification and reference of concepts can also be treated within this scheme, at least insofar as they pertain to perception. We noticed above, when talking about Austin's work, that the idea of memory and the

application of predicates played a large part in destroying the idea of certainty. But perhaps we can make clearer the line distinguishing description and identification, which Hampshire and Strawson debate ("Perception and Identification"), in terms of the cues utilized and the follow-up checks which Bruner (and before him Peirce, Dewey, and R. W. Sellars) made much of. For example, it might be possible to draw a line

for some purposes in terms of the accessibility level employed in ascribing the concepts to particular cases. Also in this area it might prove extremely fruitful to analyze the relations which obtain between the ideas of Strawson and Wittgenstein about fundamental concepts of our conceptual scheme (forms of life) and Bruner's idea of perceptual readiness and accessibility in particular.¹

The University of Chicago

Received January 15, 1969

BIBLIOGRAPHY

BOOKS

- ALEXANDER, Peter *Sensationalism and Scientific Explanation* (London, Routledge & Kegan Paul, 1963).
- ARMSTRONG, David M. *Berkeley's Theory of Vision* (Melbourne, Melbourne University Press, 1961).
- *Perception and the Physical World* (London, Routledge & Kegan Paul, 1961).
- *Bodily Sensations* (London, Routledge & Kegan Paul, 1962).
- ARNHEIM, Rudolph *Towards a Psychology of Art* (London, Faber & Faber, 1967), esp. Sects. II and III.
- ATTNEAVE, F. *Applications of Information Theory to Psychology* (New York, Holt, Reinhart & Winston, 1959).
- AUNE, Bruce *Knowledge, Mind and Nature* (New York, Random House, 1967).
- AUSTIN, John L. *Sense and Sensibilia* (Oxford, Oxford University Press, 1962). Reconstructed by G. J. Warnock.
- AYER, A. J. *The Problem of Knowledge* (London, Macmillan, 1956).
- BAKAN, Paul (ed.) *Attention* (New Jersey, Van Nostrand, 1966).
- BARTLEY, S. H. *Principles of Perception* (New York, Harper & Row, 1958).
- BELOFF, John *The Existence of Mind* (London, MacGibbon & Kee, 1962).
- BEARDSLEY, D. C. and G. V. RAMSEY (eds.) *Readings in Perception* (New Jersey, Van Nostrand, 1958).
- BERLYNE, Daniel E. *Structure and Direction in Thinking* (New York, Wiley, 1965).
- BIRREN, Faber *Color in Your World* (New York, Collier, 1962).
- BRAIN, W. Russell *The Nature of Experience* (Oxford, Oxford University Press, 1959).
- BROADBENT, D. E. *Perception and Communication* (London and New York, Pergamon, 1958).
- BRUNER, Jerome, R. R. OLIVER, and P. M. GREENFIELD, et al. *Studies in Cognitive Growth* (New York, Wiley, 1966).
- CANTRIL, H. (ed.) *The Morning Notes of Adelbert Ames, with Correspondence by John Dewey* (New Brunswick, New Jersey, Rutgers University Press, 1960).
- CHISHOLM, Roderick M. *Perceiving: A Philosophical Study*, (Ithaca, Cornell University Press, 1958).
- *Theory of Knowledge* (Englewood Cliffs, New Jersey, Prentice-Hall, 1966).
- DEMBER, William *Psychology of Perception* (New York, Holt, Reinhart & Winston, 1960).
- ECCLES, Sir John *The Brain and the Unity of Conscious Experience* (Cambridge, Cambridge University Press, 1965).
- EPSTEIN, William *Varieties of Perceptual Learning* (New York, McGraw-Hill, 1967).
- FIEANDT, Kai von *The World of Perception* (Homewood, Ill., Dorsey Press, 1966).
- FEYERABEND, Paul K., *Knowledge Without Foundations* (Oberlin, Oberlin College, 1961).
- GARNER, Wendell R. *Uncertainty and Structure as Psychological Concepts* (New York, Wiley, 1962).
- GARNETT, A. Campbell *The Perceptual Process* (London, Allen & Unwin, 1965).
- GEACH, Peter *Mental Acts* (London, Routledge & Kegan Paul, 1957).
- GEORGE, F. H. *Cognition* (London, Methuen, 1962).
- GIBSON, J. J. *Senses Considered as Perceptual Systems* (Boston, Houghton, Mifflin, 1966).
- GOMBRICH, Ernst H. *Art and Illusion* (London, Phaidon, 2nd ed., 1962).
- GREGORY, Richard L. *Eye and Brain* (London, Weidenfeld & Nicolson, 1966).

¹ I am grateful to Vere Chappell, Dudley Shapere, and Manley Thompson for discussions about an earlier draft of this essay. Also I am indebted to William Lycan, Michael McMahon, Jack Nelson, and George Schumm for discussions regarding the clarity of various parts of this essay. Of course, none of the above share any responsibility for any inaccuracy or inelegance that remains.

- CROSSMAN, Reinhardt S. *The Structure of Mind* (Madison, University of Wisconsin Press, 1965).
- GURWITSCH, Aron *The Field of Consciousness* (Pittsburgh, Duquesne University Press, 1964).
- HAMLIN, D. W. *The Psychology of Perception* (London, Routledge & Kegan Paul, 1957).
- *Sensation and Perception* (London, Routledge & Kegan Paul, 1961).
- HANSON, Norwood Russell *Patterns of Discovery* (Cambridge, Cambridge University Press, 1958).
- HARRÉ, Rom *Theories and Things* (London, Sheed & Ward, 1961).
- HAYEK, F. A. *The Sensory Order, An Inquiry into the Foundations of Theoretical Psychology* (Chicago, University of Chicago Press, 1962).
- HEISENBERG, W. *Physics and Philosophy* (London, Allen & Unwin, 1959).
- HILL, Thomas *English Contemporary Theories of Knowledge* (New York, MacMillan, 1961).
- HIRST, R. J. *The Problems of Perception* (London, Allen & Unwin, 1959).
- (ed.) *Perception and the External World* (New York, MacMillan, 1965).
- HOCHBERG, Julian E. *Perception* (Englewood Cliffs, New Jersey, Prentice-Hall, 1964).
- HOLLAND, Harry C. *The Spiral After-Illusion*, International Series of Monographs in Experimental Psychology (Oxford, Pergamon, 1965).
- ITTLESON, William H. *Visual Space Perception* (New York, Springer, 1960).
- KIDD, Aline and G. J. L. RIVOIRE *Perceptual Development in Children* (New York, International Universities Press, 1966).
- KLEE, Paul *The Thinking Eye* (New York, G. Wittenbourn, 1961) (orig., *Das Bildnerische Denken*, 1956). Trans. by R. Manheim.
- KOCH, Sigmund (ed.) *Psychology: A Study of Science*, vol. I (New York, McGraw-Hill, 1959).
- KORNER, Stephen *Conceptual Thinking* (Bristol, University of Bristol Press, 1955).
- KUHLENBECK, Hartwig *Brain and Consciousness, Some Prolegomena to an Approach to the Problem*, supplement to vol. 17, *Confinia Neurologica* (Basel, Karger, 1957).
- LANGER, Susan K. *Mind: An Essay in Feeling*, vol. I (Baltimore, Johns Hopkins University Press, 1967).
- LEIBOWITZ, Herschel W. (ed.) *Visual Perception* (New York, MacMillan, 1965).
- LOCKE, Don *Perception and Our Knowledge of the Eternal World* (London, Allen & Unwin, 1967).
- Logique et Perception*, Etudes D'Epistemologie Genetique, VI (Paris, Presses Universitaires de France, 1958).
- MANDELBAUM, Maurice *Philosophy, Science and Sense-Perception* (Baltimore, Johns Hopkins University Press, 1964).
- MERLEAU-PONTY, Maurice *The Primacy of Perception and Other Essays*, ed. by J. M. Edie (Evanston, Northwestern University Press, 1964).
- MERLEAU-PONTY, Maurice *Le Visible et l'Invisible*, ed. by C. Lefort (Paris, Gallimard, 1964).
- MOLES, Abraham *Information Theory and Esthetic Perception* (Chicago, University of Illinois Press, 1966).
- MONCRIEFF, R. W. *Odour Preferences* (New York, Wiley, 1966).
- MOORE, G. E. *Lectures on Philosophy*, ed. by C. Lewy (London, Allen & Unwin, 1966, esp. Part I).
- PASTO, T. *The Space Frame Experience in Art* (New York, A. S. Barnes, 1964).
- PAUL, L. *Persons and Perception* (London, Faber & Faber, 1961).
- PIAGET, Jean *Les Mechanismes Perceptifs: Modeles probabilistes, analyse genetique, relations avec l'intelligence* (Paris, Presses Universitaires de France, 1961).
- *Sagesse et Illusion de Philosophie* (Paris, Presses Universitaires de France, 1965).
- ROCK, Irvin *The Nature of Perceptual Adaptation* (New York, Basic Books, 1966).
- RONCHI, Vasco *Optics: The Science of Vision* (New York, New York University Press, 1957). Trans. by E. Rosen, recommended by Turbayne in his introduction to Berkeley, *Works on Vision* [New York, Library of Liberal Arts, 1963].
- SAYRE, Kenneth M. *Recognition: A Study in the Philosophy of Artificial Intelligence* (Notre Dame, University of Notre Dame Press, 1965).
- SCHEFFLER, Israel *Science and Subjectivity* (Indianapolis, Bobbs-Merrill, 1967).
- SEGALL, Marshall, Donald T. CAMPBELL and Melville J. HERSKOVITS *The Influence of Culture on Visual Perception* (New York, Bobbs-Merrill, 1966).
- SELLARS, Wilfrid *Science, Perception and Reality* (London, Routledge & Kegan Paul, 1963).
- SHOEMAKER, Sidney *Self-Knowledge and Self-Identity* (Ithaca, Cornell University Press, 1963).
- SMART, J. J. C. *Philosophy and Scientific Realism* (London, Routledge & Kegan Paul, 1963).
- SMYTHIES, J. R. *Analysis of Perception* (London, Routledge & Kegan Paul, 1956).
- SOLLEY, Charles M. and Gardner MURPHY *Development of the Perceptual World* (New York, Basic Books, 1960).
- SOLTIS, J. F. *Seeing, Knowing and Believing* (London, Allen & Unwin, 1966).
- STRAUS, Erwin *The Primary World of Senses: A Vindication of Sensory Experience* (New York, The Free Press, 1963). Trans. by J. Needleman.
- STRAWSON, P. F. *Individuals* (London, Methuen, 1959).
- *The Bounds of Sense: An Essay on Kant's Critique of Pure Reason* (London, Methuen, 1966).
- SWARTZ, Robert J. (ed.) *Perceiving, Sensing and Knowing* (New York, Doubleday-Anchor, 1965).
- TAYLOR, Charles *The Explanation of Behavior* (London, Routledge & Kegan Paul, 1964).
- TAYLOR, James G. *The Behavioral Basis of Perception* (New Haven, Yale University Press, 1962).
- TAYLOR, William *The Relationship Between Psychology and Science* (Somerset, Martigan, 1964).

- VERNON, M. D. *The Psychology of Perception* (Middlesex, Penguin, 1962).
- VERNON, M. D. (ed.) *Experiments in Visual Perception* (Middlesex, Penguin, 1966).
- VESEY, G. N. A. *The Embodied Mind* (London, Allen & Unwin, 1965).
- WALLRAFF, Charles F. *Philosophical Theory and Psychological Fact: An Attempted Synthesis* (Tucson, University of Arizona Press, 1961).
- WARNOCK, G. J. *Berkeley* (London, Penguin, 1953).
- (ed.) *The Philosophy of Perception* (Oxford, Oxford University Press, 1967).
- WHITE, Alan R. G. E. *Moore: A Critical Exposition* (Oxford, Blackwell, 1958).
- *Attention* (Oxford, Blackwell, 1964).
- WISDOM, John *Proof and Explanation*, mimeograph. Lectures given Spring 1957 at University of Virginia, prepared by S. F. Barker.
- WOLFF, Robert Paul *Kant's Theory of Mental Activity* (Cambridge, Mass., Harvard University Press, 1963).
- WOLLHEIM, Richard *On Drawing an Object* (London, H. C. Lewis, 1965).
- WYBURN, G. M., R. W. RICKFORD and R. J. HIRST *Human Senses and Perception* (Ontario, University of Toronto, 1964).
- YOLTON, John W. *Thinking and Perceiving* (La Salle, Ill., Open Court, 1962).
- ZINKERNAGEL, Peter *Conditions for Description* (London, Routledge & Kegan Paul, 1962). (Orig. 1957, trans. by O. Lindum.)
- ARTICLES
- The following abbreviations have been used:
- A: *Analysis*
 AJP: *Australasian Journal of Philosophy*
 APQ: *American Philosophical Quarterly*
 ASSP: *Aristotelian Society, Supplementary Volumes*
 BJPS: *British Journal for the Philosophy of Science*
 I: *Inquiry*
 JP: *Journal of Philosophy*
 M: *Mind*
 N: *Nous*
 PAS: *Proceedings of the Aristotelian Society*
 P: *Philosophy*
 PPR: *Philosophy and Phenomenological Research*
 PQ: *Philosophical Quarterly*
 PR: *Philosophical Review*
 PS: *Philosophy of Science*
 R: *Ratio*
 RIP: *Revue Internationale de Philosophie*
 RM: *Review of Metaphysics*
 S: *Synthese*
- AARON, R. "The Common Sense View of Sense-Perception," PAS, vol. 58 (1958), pp. 1-14.
- ACHINSTEIN, Peter "Theoretical Terms and Partial Interpretation," BJPS, vol. 14 (1964), pp. 89-105.
- ACHINSTEIN, Peter "The Problem of Theoretical Terms," APQ, vol. 2 (1965), pp. 193-203.
- ADAMS, E. M. "The Nature of the Sense-Datum Theory," M, vol. 67 (1958), 216-226.
- "Perception and the Language of Appearing," JP, vol. 55 (1958), pp. 683-690.
- "The Inadequacy of Phenomenalism," PPR, vol. 20 (1959), pp. 93-102.
- "Mind and the Language of Psychology," R, vol. 9 (1967), pp. 122-139.
- AGASSI, Joseph "Sensationalism," M, vol. 75 (1966), pp. 1-24.
- ALDRICH, V. C. "Images as Things and Things as Images," M, vol. 64 (1955), pp. 261-263.
- "Image-Mongering and Image-Management," PPR, vol. 23 (1962), pp. 51-61.
- "On Seeing Bodily Movement as Actions," APQ, vol. 4 (1967), pp. 222-230.
- ALEXANDER, Peter "Theory Construction and Theory Testing," BJPS, vol. 9 (1958), pp. 29-38.
- "Sensationalism," in P. Edwards (ed.), *The Encyclopedia of Philosophy* (New York, 1967), vol. 7.
- ALSTON, William P. "Is a Sense-Datum Language Necessary?," PS, vol. 24 (1957), pp. 41-45.
- ANSCOMBE, G. E. M. "Substance," ASSP, vol. 38 (1964), pp. 69-78.
- "The Intentionality of Sensation: A Grammatical Feature" in R. J. Butler (ed.) *Analytical Philosophy* (Oxford, 1962).
- ÅQVIST, Lennart "Notes on A. J. Ayer's 'The Terminology of Sense-Data'," A, vol. 20 (1959), pp. 106-111.
- ARDLEY, Gavin "The Nature of Perception," AJP, vol. 36 (1958), p. 189.
- ARMSTRONG, D. M. "Illusions of Sense," AJP, vol. 33 (1955), pp. 88-106.
- "A Theory of Perception" in B. B. Wolman (ed.), *Scientific Psychology: Principles and Approaches* (New York, 1965).
- ARTHADEVA, M. "Naive Realism and Illusions of Reflection," AJP, vol. 35 (1957), pp. 155-169.
- "Naive Realism and Illusions: 'The Elliptical Penny'," P, vol. 34 (1959), pp. 323-330.
- "Naive Realism and the Illusions of Refraction," AJP, vol. 37 (1959), pp. 118-137.
- "Mirror Images," AJP, vol. 38 (1960), pp. 160-162.
- ATTNEAVE, Fred "Perception and Related Areas" in S. Koch (ed.), *Psychology: A Study of Science*, vol. 4 (New York, 1962).
- ATWELL, John E. M. "Austin on Incorrigibility," PPR, vol. 27 (1966), pp. 261-266.
- AVERILL, Edward "Perception and Definition," JP, vol. 55 (1958), pp. 690-699.
- AYER, A. J. "Perception" in C. A. Mace (ed.), *British Philosophy in Mid-Century* (London, 1957).
- "Has Austin Refuted the Sense-Datum Theory?," S, vol. 17 (1967), pp. 17-40.

- BARKER, S. F. "Appearing and Appearances in Kant," *Monist*, vol. 51 (1967), pp. 426-441.
- BARNES, W. "On Seeing and Hearing" in H. D. Lewis (ed.), *Contemporary British Philosophy III* (London, 1956).
- BAYLIS, C. A. "Professor Chisholm on Perceiving," JP, vol. 56 (1959), pp. 773-791.
- "A Criticism of Lovejoy's Case for Epistemological Dualism," PPR, vol. 23 (1962), pp. 527-537.
- "Foundations for a Presentative Theory of Perception and Sensation," PAS, vol. 66 (1966), pp. 41-54.
- "Perception and Sensations as Presentational" in F. C. Dommeyer (ed.), *Current Philosophical Issues, Essays in Honor of Curt John Ducasse* (Springfield, Ill., 1966).
- BEDFORD, E. "Seeing Paints," ASSP, vol. 40 (1966), pp. 47-62. Symposium with R. Meager.
- BENNETT, Jonathan "Substance, Reality and Primary Qualities," APQ, vol. 2 (1965), pp. 1-17.
- "Real," M, vol. 75 (1966), pp. 501-515.
- BERGMANN, Gustar "Some Remarks on the Philosophy of Malebranche," RM, vol. 10 (1956), pp. 207-226.
- BERNSTEIN, Richard J. "Sellars' Vision of Man-in-the-Universe," RM, vol. 20 (1966), pp. 113-143 and 290-316.
- BERTALANFFY, L. von "An Essay on the Relativity of Categories," PS, vol. 22 (1955), 243-263.
- BERTOCCI, Peter A. "The Nature of Cognition: Minimum Requirements for a Personalistic Epistemology," RM, vol. 8 (1954), pp. 49-60.
- BLAISHARD, Brand and B. F. SKINNER "The Problem of Consciousness—A Debate," PPR, vol. 27 (1966), pp. 317-337.
- BLOCK, Irving "Truth and Error in Aristotle's Theory of Sense Perception," PQ, vol. 11 (1961), pp. 1-9.
- "On the 'Commonness' of the Common Sensibles," AJP, vol. 43 (1965), pp. 189-195.
- BLUMENFELD, David C. "On Not Seeing Double," PQ, vol. 9 (1959), pp. 264-266.
- BOLLES, R. C. and D. E. BAILEY "Importance of Object Recognition in Size Constancy," *Journal of Experimental Psychology*, vol. 51 (1956), pp. 222-225.
- BOUWSMA, O. K. "Reflections on Moore's Latest Book," PR, vol. 64 (1955), pp. 248-263. (*On Some Main Problems.*)
- BOWER, T. R. G. "The Visual World of Infants," *Scientific American*, vol. 216 (1967), pp. 80-92.
- BRADLEY, R. D. "Avowals of Immediate Experience," M, vol. 73 (1964), pp. 186-203.
- BRAIN, Russell "Space and Sense-Data," BJPS, vol. 11 (1960), pp. 177-191.
- BROAD, C. D. "Reply to My Critics" in P. A. Schilpp (ed.), *The Philosophy of C. D. Broad* (Tudor, 1959).
- BRONAUGH, Richard N. "The Argument from the Elliptical Penny," PQ, vol. 14 (1964), pp. 151-157.
- BROWN, Norman "Sense-data and Material Objects," M, vol. 66 (1957), pp. 173-194.
- BRUNER, Jerome and L. MINTURN "Perceptual Identification and Perceptual Organization," *Journal of General Psychology*, vol. 53 (1955), pp. 21-28.
- "On Perceptual Readiness," *Psychological Review*, vol. 64 (1957), pp. 123-152, reprinted in R. J. C. Harper, C. C. Anderson, G. M. Christensen and S. M. Hunka (eds.), *The Cognitive Processes* (Englewood Cliffs, New Jersey, 1964).
- "Les processus de Preparation a la Perception" in *Logique et Perception, Etudes D'Epistemologie Genetique*, vol. VI (Paris, 1958).
- BUCKLEW, John "The Subjective Tradition in Phenomenological Psychology," PS, vol. 22 (1955), pp. 289-299.
- BURGENER, R. J. C. "Price's Theory of the Concept," RM, vol. 11 (1957), pp. 143-159.
- CALHOUN, Edward "Human Likeness and the Formation of Empirical Concepts," RM, vol. 13 (1959), pp. 383-395.
- CAMPBELL, Donald T. "Methodological Suggestions from a Comparative Psychology of Knowledge Processes," I, vol. 2 (1959), pp. 152-182.
- CAPEK, Milec "The Development of Reichenbach's Epistemology," RM, vol. 11 (1957), pp. 42-67.
- CHARI, C. T. K. "On the 'Space' and 'Time' of Hallucinations," BJPS, vol. 8 (1957), pp. 302-306.
- CHILD, Arthur "Projection," P, vol. 42 (1967), pp. 20-36.
- CHISHOLM, Roderick M. "Epistemic Statements and the Ethics of Belief," PRR, vol. 16 (1955), pp. 447-460.
- "Appear,' 'Take,' and 'Evident'," JP, vol. 53 (1956), pp. 729-739, represented in Swartz.
- "Evidence as Justification," JP, vol. 58 (1961), pp. 739-748.
- "The Principles of Epistemic Appraisal" in F. C. Dommeyer (ed.), *Current Philosophical Issues, Essays in Honor of Curt John Ducasse* (Springfield, Ill., 1966).
- "On Some Psychological Concepts and the 'Logic' of Intentionality" in H. N. Castaneda (ed.), *Intention, Mind, and Perception* (Detroit, 1967). Comments by R. Sleight.
- and Wilfrid SELLAISIE "Intentionality and the Mental" in H. Feigl, M. Scriven and G. Maxwell (eds.), *Minnesota Studies in the Philosophy of Science*, vol. II (Minneapolis, 1958).
- CHOMSKY, Noam "Perception and Language," summary in M. Wartofsky (ed.), *Boston Studies in the Philosophy of Science*, vol. I (Dordrecht-Holland, 1963).
- CLARK, Michael "Intentional Objects," A, vol. 25 (1964), pp. 123-131.
- "Knowledge and Grounds," A, vol. 24 (1963), pp. 46-48.
- CLARKE, Thompson "Seeing Surfaces and Physical Objects" in M. Black (ed.), *Philosophy in America* (Ithaca, 1965).
- CLEMENT, W. C. "Seeing and Hearing," BJPS, vol. 6 (1955), pp. 61-63.
- "Quality Orders," M, vol. 65 (1956), pp. 184-199. Cf. letter from N. Goodman (M, vol. 66 [1957], p. 78).

- CLIFFORD, Paul R. "Knowledge as Trans-sensational," *RM*, vol. 7 (1963), pp. 361-371.
- COHEN, A. "On Methods in the Analysis of Speech Perception," *XYZ*, vol. 517 (1967), pp. 331-343.
- COLLINS, Arthur W. "The Epistemological Status of the Concept of Perception," *PR*, vol. 76 (1967), pp. 436-459.
- COPPLESTON, F. C. "On Seeing and Noticing," *P*, vol. 29 (1954), pp. 152-157.
- COUSIN, D. R. "Naive Realism," *PAS*, vol. 55 (1955), pp. 179-200.
- COX, J. W. Roxbee "Are Perceptible Qualities 'in' Things?," *A*, vol. 23 (1962), pp. 97-103.
- CUMMING, Philip "Perceptual Relativity and Ideas in the Mind," *PPR*, vol. 24 (1963), pp. 202-214.
- DANTO, Arthur C. "Concerning Mental Pictures," *JP*, vol. 55 (1958), pp. 12-20.
- DANIELS, Charles B. "Colors and Sensations: Or How to Define a Pain Ostensively," *APQ*, vol. 4 (1967), pp. 231-237.
- DAY, J. P. "Unconscious Perception," *ASSP*, vol. 34 (1960), pp. 47-66. Symposium with Vesey.
- DEUTSCHER, Max "David Armstrong and Perception," *AJP*, vol. 41 (1963), pp. 81-88 (reply by Armstrong, p. 246).
- DORE, Clement "Ayer on the Causal Theory of Perception," *M*, vol. 73 (1964), pp. 287-290.
- "Seeming to See," *APQ*, vol. 2 (1965), pp. 312-318.
- DRETSKE, Fred "Observational Terms," *PR*, vol. 73 (1965), pp. 25-42.
- "Ziring Ziderata," *M*, vol. 75 (1966), pp. 211-223.
- DREYFUS, H. L. and S. J. TODD "The Three Worlds of Merleau-Ponty," *PPR*, vol. 22 (1961), pp. 559-565.
- DUCASSE, Curt J. "Causation: Perceivable? Or Only Inferred?," *PPR*, vol. 26 (1965), pp. 173-179.
- "Minds, Matter and Bodies" in J. R. Smythies (ed.), *Brain and Mind* (London, 1965). Comments by Brain, Flew, Price and Smythies.
- "How Literally Causation is Perceivable?," *PPR*, vol. 28 (1967), pp. 271-273.
- DUGGAN, T. "Thomas Reid's Theory of Sensation," *PR*, vol. 69 (1960), pp. 90-100.
- DUNCKER, Karl "Phenomenology and Epistemology of Consciousness of Objects," *PPR*, vol. 7 (1946), pp. 505-541; trans. by Luise Haessler.
- EBERSOLE, Frank B. "On Seeing Things," *PQ*, vol. 11 (1961), pp. 289-300.
- "How Philosophers See Stars," *M*, vol. 74 (1965), pp. 509-529.
- ELLIS, Brian "Physical Monism," *S*, vol. 17 (1967), pp. 141-161.
- ELLISON, L. G. "The Scientists' Criterion of True Observation," *PS*, vol. 30 (1963), pp. 41-52.
- EPSTEIN, Joseph "Professor Ayer on Sense-Data," *JP*, vol. 53 (1956), pp. 401-415.
- FARBERG, M. and T. NORDENSTAM "If I Carefully Examine a Visual After-Image, What Am I Looking At, and Where Is It?," *A*, vol. 19 (1958), pp. 99-100.
- FEIGL, Herbert "Philosophical Embarrassments of Psychology," *American Psychologist*, vol. 14 (1959), p. 115.
- FEYERABEND, P. K. "An Attempt at a Realistic Interpretation of Experience," *PAS*, vol. 58 (1958), pp. 143-170.
- "Patterns of Discovery," *PR*, vol. 69 (1960), pp. 247-252.
- FIRTH, Roderick "Ultimate Evidence," *JP*, vol. 53 (1956), pp. 732-739. (Represented in Swartz.)
- "Chisholm and the Ethics of Belief," *PR*, vol. 68 (1959), pp. 493-506.
- "Austin and the Argument from Illusion," *PR*, vol. 72 (1964), pp. 27-28.
- "Coherence, Certainty, and Epistemic Priority," *JP*, vol. 61 (1964), pp. 545-557.
- "The Anatomy of Certainty," *PR*, vol. 76 (1967), pp. 3-27.
- "The Men Themselves: Or the Role of Causation in Our Concept of Seeing" in H.-N. Castaneda (ed.), *Intentionality, Minds and Perception* (Detroit, 1967). Comments by Chas. Caton.
- FLEMING, B. Noel "Recognizing and Seeing As," *PR*, vol. 66 (1967), pp. 161-179.
- "The Nature of Perception," *RM*, vol. 16 (1962), pp. 259-295.
- "The Idea of a Solid," *AJP*, vol. 43 (1965), pp. 131-143.
- "Price on Infallibility," *M*, vol. 75 (1966), pp. 193-210.
- FODOR, J. A. "Could There Be a Theory of Perception?," *JP*, vol. 63 (1966), pp. 369-380.
- FRENCH, David "The Relation of Anthropology to Studies in Perception and Cognition" in S. Koch (ed.), *Psychology: A Study in Science*, vol. 6 (New York, 1963).
- FRITZ, Charles A. "Sense Perception and Material Objects," *PPR*, vol. 16 (1955), pp. 303-316.
- "Contextual Properties and Perception," *PPR*, vol. 20 (1959), pp. 338-351.
- FURLONG, E. J. "Berkeley and the 'Knot About Inverted Images'," *AJP*, vol. 41 (1963), pp. 306-316.
- GAFFRON, M. "Some New Dimensions in the Phenomenal Analysis of Visual Experiences," *Journal of Personality*, vol. 24 (1956), pp. 285-307.
- GARDNER, R. W., H. W. HAKE and C. W. ERIKSON "Operationism and the Concept of Perception," *Psychological Review*, vol. 63 (1956), pp. 149-159.
- "Cognitive Controls of Attention Deployment as Determinants of Visual Illusions," *Journal of Abnormal Sociology and Psychology*, vol. 62 (1961), pp. 120-127.
- GAULD, Alan "Could a Machine Perceive?," *BJPS*, vol. 17 (1966), pp. 44-58.
- GEORGE, F. H. "Epistemology and the Problem of Perception," *M*, vol. 66 (1957), pp. 491-506.
- GIBSON, E. and R. WALK "The 'Visual Cliff'," *Scientific American*, vol. 202 (1960), p. 64.
- GIBSON, J. J. and E. J. GIBSON "Perceptual Learning: Differentiation or Enrichment?," *Psychological Review*, vol. 62 (1955), pp. 32-41.

- GIBSON, James J. "New Reasons for Realism," S, vol. 17 (1967), pp. 162-172.
- GIBSON, Q. "Is There a Problem about Appearances?" PQ, vol. 16 (1966), pp. 319-328.
- GOODMAN, Nelson "The Way the World Is," RM, vol. 14 (1960), pp. 48-56.
- "Review of Gombrich's *Art and Illusion*," JP, vol. 57 (1960), pp. 595-599.
- GRAHAM, C. H. "Sensation and Perception in an Objective Psychology," *Psychological Review*, vol. 65 (1958), pp. 65-76.
- and Phillburn RATOOSH, "Notes on Some Interrelations of Sensory Psychology, Perception, and Behavior" in S. Koch (ed.), *Psychology: A Study of Science*, vol. 4 (New York, 1962).
- GREGORY, R. L. "Seeing in Space," *Cambridge Research*, vol. 7 (1965), pp. 5-7.
- GRICE, H. P. "The Causal Theory of Perception," ASSP, vol. 35 (1961), pp. 121-152. Symposium with A. R. White. Represented in Swartz and in Warnock.
- "Some Remarks About the Senses" in R. J. Butler (ed.), *Analytical Philosophy* (Oxford, 1962).
- GRIFFITHS, A. Phillips "Ayer on Perception," M, vol. 60 (1960), pp. 486-498.
- GROSSMAN, Reinhardt "Sensory Intuition and the Dogma of Localization," I, vol. 5 (1962), pp. 238-251.
- GURWITSCH, Aron "Contribution to the Phenomenological Theory of Perception" in *Studies in Phenomenology and Psychology* (Evanston, Ill., 1966).
- "On a Perceptual Root of Abstraction," *ibid.*
- HALL, E. W. "The Adequacy of a Neurological Theory of Perception," PPR, vol. 20 (1959), pp. 75-84.
- HALL, R. "The Term 'Sense Datum'," M, vol. 73 (1964), pp. 130-131.
- HALSBURY, Earl of "Epistemology and Communication Theory," P, vol. 34 (1959), pp. 289-307.
- HAMLIN, D. W. "Psychological Explanation and the Gestalt Hypothesis," M, vol. 60 (1951), pp. 506-520.
- "The Visual Field and Perception," ASSP, vol. 31 (1957), pp. 107-124. Symposium with A. C. Lloyd.
- HAMPSHIRE, Stuart "Identification and Existence" in H. K. Lewis (ed.), *Contemporary British Philosophy III* (London, 1956).
- "Perception and Identification," ASSP, vol. 35 (1961), pp. 81-96. Symposium with P. F. Strawson.
- HANSON, Norwood Russell "On Having the Same Visual Experiences," M, vol. 69 (1960), pp. 340-350.
- "Observation and Interpretation" in *Voice of America, Forum Lectures*, Philosophy of Science Series, No. 9, 1964. Reprinted in S. Morgenbesser (ed.), *Philosophy of Science Today* (New York, 1967).
- HARDIE, W. F. R. "Ordinary Language and Perception," PQ, vol. 5 (1955), pp. 97-108.
- "Austin on Perception," P, vol. 38 (1963), pp. 253-263.
- HARRIS, Errol "The Mind-Dependence of Objects," PQ, vol. 6 (1956), pp. 223-235.
- HARRISON, Bernard "On Describing Colours," I, vol. 10 (1967), pp. 38-52.
- HARROD, Sir Roy "Sense and Sensibilia," P, vol. 38 (1963), pp. 227-242.
- HARTLAND-SWANN, John "Being Aware of" and 'Knowing', PQ, vol. 7 (1957), pp. 126-135.
- "On Describing the World," AJP, vol. 34 (1956), pp. 106-117.
- HARTNACK, J. "Remarks on the Concept of Sensation," JP, vol. 56 (1959), pp. 111-117.
- HARTSHORNE, Charles "Professor Hall on Perception," PPR, vol. 21 (1960), pp. 563-571.
- HELD, R. and A. HEIN "Movement Produced Stimulation in the Development of Visually Guided Behavior," *Journal of Comparative Psychology*, vol. 56 (1963), pp. 872-876.
- HEMPEL, C. G. "The Theoretician's Dilemma" in H. Feigl, M. Scriven and G. Maxwell (eds.), *Minnesota Studies in the Philosophy of Science*, vol. II (Minneapolis, 1958), p. 158.
- HESSE, Mary "Theories, Dictionaries, and Observation," BJPS, vol. 9 (1958), pp. 12-28. Cf. reply by Alexander; note by Hesse, p. 128.
- HINSHAW, Virgil G. "The Given," PPR, vol. 18 (1957), p. 312.
- HINTON, J. M. "Seeming and Causes," P, vol. 41 (1966), pp. 348-355.
- "Visual Experiences," M, vol. 76 (1967), pp. 217-227.
- "Experiences," PQ, vol. 17 (1967), pp. 1-13.
- "Illusions and Identity," A, vol. 27 (1967), pp. 67-76.
- "On Not Having What You Are Given," I, vol. 10 (1967), pp. 313-316.
- "Perception and Identification," PR, vol. 76 (1967), pp. 421-435.
- HIRST, R. J. "The Difference Between Sensing and Observing," ASSP, vol. 28 (1959), pp. 197-218. Symposium with R. Wollheim, reprinted in Warnock.
- "Critical Study of Austin's *Sense and Sensibilia*," PQ, vol. 13 (1963), pp. 162-170.
- "Form and Sensation," ASSP, vol. 39 (1965), pp. 155-172. Symposium with C. J. F. Williams.
- "Perception" in P. Edwards (ed.), *The Encyclopedia of Philosophy*, vol. 6 (New York, 1967).
- "Phenomenalism," *ibid.*
- "Sensa," *ibid.*, vol. 7.
- HOCHBERG, Herbert "Ontology and Acquaintance," *Philosophical Studies*, vol. 17 (1966), pp. 49-55.
- HOCHBERG, T. E. "Nativism and Empiricism in Perception" in Leo Postman (ed.), *Psychology in the Making* (New York, 1962).
- HOFFMAN, Robert "Mr. Malinovich on 'Seeing As' As An Achievement," PPR, vol. 27 (1966), pp. 439-440.
- HOROWITZ, Mardi J. "Visual Imagery and Cognitive Organization," *American Journal of Psychiatry*, vol. 123 (1967), p. 938.
- HUBEL, David H. "The Visual Cortex of the Brain," *Scientific American*, vol. 209 (1963), pp. 54-62.

- HUDSON, H. "Achievement Expression," A, vol. 16 (1955), pp. 127-130.
- HUTCHINGS, P. A. E. "What is a Proper Usage of 'Illusion'?" AJP, vol. 34 (1956), pp. 38-42.
- INGRAM-PEARSON, E. W. "The Reality of Appearances," RM, vol. 9 (1955), pp. 200-206.
- ITTELSON, William H. "Perception and Transactional Psychology" in *Psychology: A Study of Science*, vol. 4 (New York, 1962).
- JENKIN, N. "Two Types of Perceptual Experience," *Journal of Clinical Psychology*, vol. 12 (1956), pp. 44-48.
- JONCKHEERE, A. "Geometrie et Perception" in *La Lecture de l'Experience, Etudes D'Epistemologie Genetique*, V (Paris, 1958).
- JOSKE, W. D. "Inferring and Perceiving," PR, vol. 72 (1963), pp. 433-445.
- KAAAL, Hans "Senses of 'Perceive' or Senses of 'Senses of Perceive'," A, vol. 24 (1963), pp. 6-11.
- KAMINSKY, Jack "Dewey's Concept of An Experience," PPR, vol. 17 (1956), pp. 316-330.
- KENNY, Anthony "The Argument From Illusion in Aristotle's Metaphysics Gamma (1009-10)," M, vol. 76 (1967), pp. 184-197.
- KHATCHADOURIAN, Haig "About Imaginary Objects," R, vol. 8 (1966), pp. 77-89.
- KING-FARLOW, John "Senses and Sensibilia," A, vol. 23 (1962), pp. 37-40.
- KNEALE, W. C. "What Can We See?" in S. Korner (ed.), *Observation and Interpretation in the Philosophy of Physics, 1957* (New York, 1962).
- KNOX, John, Jr. "Concerning the Argument from Perspectival Variation," RM, vol. 15 (1961), pp. 518-521.
- "On Mr. Nelson's Refutation of Sense-Data," R, vol. 8 (1966), p. 90.
- "The Logic of Appearing," I, vol. 10 (1967), pp. 245-250.
- KOTARBINSKA, Janina "On Ostensive Definitions," PS, vol. 27 (1960), pp. 1-22.
- KULLMAN, Michael and C. TAYLOR "The Pre-Objective World," RM, vol. 12 (1958), pp. 108-132.
- LANGFORD, C. H. and Marion "Appearances and Reality in Perception," PPR, vol. 20 (1959), pp. 532-534.
- LAZEROWITZ, Morris "Austin's *Sense and Sensibilia*," P, vol. 38 (1963), pp. 242-252.
- LEE, Donald S. "The Construction of Empirical Concepts," PPR, vol. 27 (1966), pp. 183-198.
- LEMONS, Ramon M. "Immediacy, Privacy and Ineffability," PPR, vol. 25 (1961), pp. 500-515.
- "Sensation, Perception, and the Given," R, vol. 6 (1964), pp. 63-80.
- LENNEBERG, Erich "The Relation of Language to the Formation of Concepts," S, vol. 14 (1962), pp. 103-109.
- LETTYIN, J. Y., H. R. MATURANA, W. S. MCCULLOCH and W. H. PITTS "What the Frog's Eye Tells the Frog's Brain," *Proceedings of the IRE*, vol. 47 (1959), pp. 1940-1959, reprinted in W. McCulloch *Embodiments of Mind* (Cambridge, Mass., 1965).
- LEVINAS, Emmanuel "Intentionalite et Sensation," RIP, vol. 19 (1965), p.p. 34-54.
- LEVY, Erwin "On the Possibility of a Perceptual World-in-Common," PPR, vol. 28 (1967), pp. 48-57.
- LEWIS, C. I. "Realism or Phenomenalism," PR, vol. 64 (1955), pp. 233-247.
- LEWIS, David K. "Percepts and Color Mosaics in Visual Experience," PR, vol. 75 (1966), pp. 357-368.
- LOYD, A. C. "The Visual Field and Perception," ASSP, vol. 31 (1957), pp. 125-144. Symposium with D. W. Hamlyn.
- LOCKE, Don "Appearance-Determined Qualities," A, vol. 28 (1967), pp. 39-42.
- LYCOS, K. "Images and the Imaginary," AJP, vol. 43 (1965), pp. 321-338.
- MACIVER, A. M. "Knowledge," ASSP, vol. 32 (1958), pp. 1-24.
- MADDEN, Edward H. "E. G. Boring's Philosophy of Science," PS, vol. 32 (1965), pp. 194-201.
- MALINOVICH, Stanley "Perception: An Experience or An Achievement," PPR, vol. 25 (1964), pp. 161-168.
- MANDELBAUM, Maurice "Definiteness and Coherence in Sense-Perception," N, vol. 1 (1967), pp. 123-138.
- MARC-WOGAU, Konrad "On C. D. Broad's Theory of Senses" in P. A. Schilpp (ed.), *The Philosophy of C. D. Broad* (Tudor, 1959).
- "Gilbert Ryle on Sensation" in *Philosophical Essays Dedicated to Gunnar Aspelin* (Lund, 1963).
- MARGOLIS, Joseph "If I Carefully Examine a Visual After-Image, What Am I Looking At and Where Is It?" A, vol. 19 (1958), pp. 97-99.
- "Nothing Can Be Heard But Sound," A, vol. 20 (1959), pp. 82-87.
- "How Do We Know That Anything Continues to Exist When It Is Unperceived?" A, vol. 21 (1960), pp. 105-108.
- "Certainty About Sensations," PPR, vol. 25 (1964), p. 242.
- "After-Images and Pains," P, vol. 41 (1966), pp. 333-340.
- "Perception, Inferences, and Mediation," JP, vol. 64 (1967), pp. 119-123.
- "Awareness of Sensations and of the Location of Sensations," A, vol. 27 (1966), pp. 29-32.
- MASCALL, E. L. "Perception and Sensation," PAS, vol. 64 (1964), pp. 259-272.
- MATES, Benson "Sense-Data," I, vol. 10 (1967), pp. 225-244.
- MAXWELL, Grover "The Ontological Status of Theoretical Entities" in H. Feigl and G. Maxwell (eds.), *Minnesota Studies in the Philosophy of Science*, vol. III (Minneapolis, 1962).
- MAXWELL, Nicholas "Physics and Common Sense," BJPS, vol. 16 (1966), pp. 295-311.
- MCCULLOCH, Warren S. "A Historical Introduction to Postulation Foundations in Experimental Epistemo-

- logy" in F. S. C. Northrop and H. H. Livingston (eds.), *Cross-Cultural Understanding* (New York, 1965). Reprinted in McCulloch, *Embodiments of Mind* (Cambridge, Mass., 1965).
- MCDANIEL, S. V. "A Note on the Percept Theory," *M*, vol. 72 (1963), pp. 409-413.
- MCGILL, U. J. "Epistemological Dualism and the Partition," *PPR*, vol. 23 (1962), pp. 511-526.
- McKINNEY, J. P. "Phenomenalism: A Survey and Re-assessment," *AJP*, vol. 37 (1959), pp. 221-233.
- McLENDON, Hiram J. "Has Russell Proved Naive Realism Self-Contradictory?," *JP*, vol. 53 (1957), pp. 289-302.
- MEAGER, Ruby "Seeing Paintings," *ASSP*, vol. 40 (1966), pp. 63-89. Symposium with E. Bedford.
- MELLOR, W. W. "The Incorrigible," *PQ*, vol. 15 (1965), pp. 35-42.
- MICHOTTE, A. "Perception and Cognition," *Acta Psychol.*, vol. I (1957), p. 13.
- MOORE, Asher "Chisholm on Intentionality," *PPR*, vol. 21 (1960), pp. 248-254.
- MOORE, G. E. "Visual Sense Data" in C. A. Mace (ed.), *British Philosophy in the Mid-Century* (London, 1957). Reprinted in Swartz.
- *Commonplace Book*, ed. by C. Lewy (London, 1962). I: pp. 9, 10, 25, 27, 29, 30, 32, 34, 35. III: pp. 2, 4, 10, 11, 13, 15, 17-20. IV: pp. 1, 2, 4, 10, 14, 16. V: pp. 4, 8, 11, 12, 17, 20, 21, 23, 25. VI: p. 10. VII: p. 11. IX: p. 17.
- MORGENBESSER, Sidney "Perception: Cause and Achievement," summary in M. Wartofsky (ed.), *Boston Studies in the Philosophy of Science*, vol. I (Dordrecht-Holland, 1963).
- MOURELATOS, Alexander "The Real Appearances, and Human Error in Early Greek Philosophy," *RM*, vol. 19 (1965), pp. 346-365.
- MUNDLE, C. W. K. "Common Sense versus Mr. Hirst's Theory of Perception," *PAS*, 60 (1960), pp. 61-78 (followed by a reply by Hirst).
- "Primary and Secondary Qualities," *A*, vol. 28 (1967), pp. 33-38.
- MURPHY, G. "Affect and Perceptual Learning," *Psychological Review*, vol. 63 (1956), pp. 1-15.
- MYERS, Charles Mason "On Actually Seeing," *Philosophical Studies*, vol. 8 (1957), pp. 28-32.
- "Phenomenological Idiom and Perceptual Mode," *PS*, vol. 28 (1958), pp. 71-81.
- "The Determinate and Determinable Modes of Appearing," *M*, vol. 67 (1958), pp. 32-49.
- "Phenomenal Organization and Perceptual Mode," *P*, vol. 34 (1959), pp. 331-337.
- "Perceptual Events, States, and Processes," *PS*, vol. 29 (1963), pp. 285-291.
- MYERS, G. E. "Perception and the 'Time-lag' Argument," *A*, vol. 17 (1956), pp. 97-102.
- "Perception and the Sentience Hypothesis," *M*, vol. 72 (1963), pp. III-20.
- NAESS, Arne and Siri "Psychological Research and Human Problems," *PS*, vol. 27 (1960), pp. 134-146.
- NAGEL, Ernest "Review of Hutten's *The Language of Modern Physics*," *BJPS*, vol. 10 (1959), p. 249.
- NAKHNIKIAN, George "Plato's Theory of Sensation," *RM*, vol. 9 (1955), pp. 129-148 and 306-327.
- "A Note on Plato's Theory of Sensation," *RM*, vol. 10 (1956), pp. 355-356.
- NATSOULAS, Thomas "What Are Perceptual Reports About?," *Psychological Bulletin*, vol. 67 (1967), pp. 249-272.
- NELSON, John O. "On the Impossibility of Sense Data," *R*, vol. 6 (1964), pp. 145-160.
- "An Examination of D. M. Armstrong's Theory of Perception," *APQ*, vol. 1 (1964), pp. 154-160.
- "Tastes," *PPR*, vol. 26 (1966), pp. 537-545.
- ODEGARD, Douglas "Sensations as Qualities," *PQ*, vol. 17 (1967), pp. 308-316.
- OLIVER, W. Donald "The Logic of Perspective Realism," *JP*, vol. 35 (1938), pp. 197-208.
- O'SHAUGHNESSY, Brian "The Location of Sound," *M*, vol. 66 (1957), pp. 471-490.
- "An Impossible Auditory Experience," *PAS*, vol. 57 (1957), pp. 53-82.
- "Material Objects and Perceptual Standpoints," *PAS*, vol. 65 (1965), pp. 77-98.
- PASTORE, Nicholas "Condillac's Phenomenological Rejection of Locke and Berkeley," *PPR*, vol. 27 (1967), pp. 429-431.
- PETERS, R. "Observationalism in Psychology," *M*, vol. 60 (1951), pp. 43-61.
- PFAFFMANN, Carl "Sensory Processes and their Relation to Behavior: Studies on the Senses of Taste as a Model S-R System" in S. Koch (ed.), *Psychology: A Study of Science*, vol. 4 (New York, 1962).
- PHILLIPS, Robert L. "Austin and Berkeley, On Perception," *P*, vol. 39 (1964), pp. 161-163.
- PIAGET, Jean "Assimilation et Connaissance" in *La Lecture de l'expérience, Etudes D'Epistemologie Genetique*, V (Paris, 1958).
- and A. MORF "Les Isomorphismes Partiels entre les Structures Logiques et les Structures Perceptives" in *Logique et Perception, Etudes D'Epistemologie Genetique*, VI (Paris, 1958).
- "Les 'Preinferences' Perceptives et leurs Relations avec les Schemes Sensori-moteurs et Operatures," *ibid.*
- PIKE, Alfred "The Phenomenological Approach to Musical Perception," *PPR*, vol. 27 (1966), pp. 247-254.
- PLACE, U. T. "Consciousness and Perception in Psychology," *ASSP*, vol. 40 (1966), pp. 101-124. Symposium with A. Watson.
- PLAUT, N. C. "Empiricism, Solipsism and Realism," *BJPS*, vol. 18 (1962), pp. 216-228.
- POLANYI, Michael "Sense-Giving and Sense-Reading," *P*, vol. 42 (1967), pp. 301-325.

- POLLOCK, John L. "Criteria and Our Knowledge of the Material World," *PR*, vol. 76 (1967), pp. 28-60.
- PRICE, H. H. "The Argument From Illusion" in H. D. Lewis (ed.), *Contemporary British Philosophy III* (London, 1956).
- "Professor Ayer on the Problem of Knowledge," *M*, vol. 67 (1958), pp. 433-464.
- "Comment on Burgener," *RM*, vol. 12 (1958), pp. 481-485.
- "Sir Russell Brain on the Modes of Apprehension," *BJPS*, vol. 11 (1960), pp. 71-76.
- "The Nature and the Status of Sense-Data in Broad's Epistemology" in P. A. Schilpp (ed.), *The Philosophy of C. D. Broad* (Tudor, 1959).
- "Appearing and Appearances," *APQ*, vol. 1 (1964), pp. 3-19.
- PRONKO, N. H., R. EBERT and G. GREENBERG "A Critical Review of Theories in Perception" in A. Kidd and J. Rivoire (eds.), *Perceptual-Development in Children* (New York, 1966).
- PUSTILNIK, Jack "Austin's Epistemology and His Critics," *P*, vol. 39 (1964), pp. 163-173.
- QUINTON, A. "The Problem of Perception," *M*, vol. 64 (1955), pp. 28-51. Reprinted in Swartz and in Warnock.
- "Matter and Space," *M*, vol. 73 (1964), pp. 332-352.
- "The Foundations of Knowledge" in B. Williams and A. Montofiore (eds.), *British Analytical Philosophy* (New York, 1966).
- RANKEN, Nani L. "A Note on Ducasse's Perceivable Causation," *PPR*, vol. 28 (1967), pp. 269-270.
- RANKIN, K. W. "Ayer's Anti-Phenomenalism," *AJP* vol. 36 (1958), pp. 109-119.
- RATLIFF, Floyd "Some Interrelations among Physics, Physiology, and Psychology in the study of Vision" in S. Koch (ed.), *Psychology: A Study of Science*, vol. IV (New York, 1962).
- RESCHER, Nicholas "Presuppositions of Knowledge," *RIP*, vol. 13 (1959), pp. 418-429.
- and Paul OPPENHEIM "Logical Analysis of Gestalt Concepts," *BJPS*, vol. 6 (1955), pp. 89-106.
- RESNICK, Lawrence "Empiricism and Natural Kinds," *JP*, vol. 57 (1960), p. 559.
- RICHMAN, Robert J. "The Whereabouts of Percepts," *JP*, vol. 55 (1958), pp. 344-348.
- RITCHIE, D. M. "Can Animals See? A Cartesian Query?" *PAS*, vol. 64 (1964), pp. 221-242.
- ROBERTS, Fred S. and Patrick SUPPES "Some Problems in the Geometry of Visual Perceptions," *S*, vol. 17 (1967), pp. 173-201.
- ROLSTON, Howard L. "Kinaesthetic Sensations Revisited," *JP*, vol. 62 (1965), pp. 96-100.
- ROMMETVEIT, Ragnar "Epistemological Notes on Recent Studies of Social Perception," *I*, vol. 1 (1958), pp. 213-231.
- ROSS, J. J. "The Reification of Appearance," *P*, vol. 40 (1965), pp. 113-128.
- RYLE, Gilbert "Sensation" in H. D. Lewis (ed.), *Contemporary British Philosophy* (London, 1956). Reprinted in Schwartz.
- SANDLE, Douglas "The Science of Art," *Science Journal*, vol. 3 (1967), pp. 80-85.
- SAYRE, Kenneth M. "On Disagreements About Perception," *I*, vol. 7 (1964), pp. 143-162.
- SCHLAGEL, Richard H. "Language and Perception," *PPR*, vol. 23 (1962), pp. 192-204.
- SCHMITT, Richard "Maurice Merleau-Ponty I and II," *RM*, vol. 19 (1965), pp. 492-516 and 782-741.
- SCRIVEN, Michael "Modern Experiments in Telepathy," *PR*, vol. 65 (1956), pp. 231-253.
- SELLARS, R. W. "Sensations as Guides to Perceiving," *M*, vol. 68 (1959), pp. 2-15.
- SELLARS, Wilfrid "Empiricism and the Philosophy of Mind" in H. Feigl and M. Scriven (eds.), *Minnesota Studies in the Philosophy of Science*, vol. I (Minneapolis, 1956). Reprinted in *Science, Perception and Reality* (London, 1963).
- "Intentionality and the Mental" in H. Feigl, M. Scriven and G. Maxwell (eds.), *Minnesota Studies in the Philosophy of Science*, vol. II (Minneapolis, 1958). With R. Chisholm.
- "The Language of Theories" in H. Feigl and G. Maxwell (eds.), *Current Issues in the Philosophy of Science* (New York, 1961). Reprinted in *Science, Perception and Reality* (London, 1963).
- "Phenomenalism" in *Science, Perception and Reality* (London, 1963). Also in H. N. Castaneda, *Intentionality, Minds and Perception* (Detroit, 1967). Comments by Bruce Aune.
- "A Refutation of Phenomenalism: Prolegomena to a Defense of Scientific Realism" in P. Feyerabend and G. Maxwell (eds.), *Mind, Matter, and Method. Essays in Honor of Herbert Feigl* (Minneapolis, 1966).
- "Some Reflections on Thoughts and Things," *N*, vol. 1 (1967), pp. 97-121.
- SEVERENS, Richard "Seeing," *PPR*, vol. 28 (1967), pp. 213-221.
- SIBLEY, F. N. "Seeking, Scrutinizing and Seeing," *M*, vol. 64 (1955), pp. 455-478. Reprinted in Warnock.
- SIEGLER, F. A. "Probability, Certainty, and Illusions," *I*, vol. 5 (1962), pp. 91-115.
- SIRGELLO, G. "Perceptual Acts and Pictorial Art: A Defense of the Expression Theory," *JP*, vol. 62 (1965), pp. 669-677.
- SLAKELY, Thomas "Aristotle on 'Sense Perception'," *PR*, vol. 70 (1961), pp. 470-489.
- SMART, J. J. C. "Colours," *P*, vol. 36 (1961), pp. 128-142.
- "Physical Objects and Physical Theories" in *Philosophy and Scientific Realism* (London, 1963).
- "The Secondary Qualities," *ibid*.
- SMULLYAN, Arthur "Aspects," *PR*, vol. 64 (1955), pp. 33-42.
- SMYTHIES, J. R. "The Stroboscopes as Providing

- Empirical Confirmation of the Representative Theory of Perception," BJPS, vol. 6 (1955), pp. 332-335.
- "A Note on Martin Lean's *Sense-Perception and Matter*," *Philosophical Studies*, vol. 6 (1955), pp. 4-8.
- "On Some Properties and Relations of Images," PR, vol. 67 (1958), pp. 389-394.
- "On Space and Time of Images," BJPS, vol. 9 (1958), pp. 40-42.
- "'Philosophical' and 'Scientific' Sense-Data," BJPS, vol. 9 (1958), pp. 224-225.
- "The Problem of Perception," BJPS, vol. 11 (1960), pp. 224-238.
- "Some Recent Theories of Mind" in I. T. Ramsey (ed.) *Biology and Purpose* (Oxford, 1965). Comments by H. H. Price and A. M. Quinton.
- "The Representative Theory of Perception" in J. R. Smythies (ed.), *Brain and Mind* (London, 1965). Comments by R. Brain and H. H. Price.
- SPECTOR, Marshall "Theory and Observation, I and II," BJPS, vol. 17 (1966), pp. 1-20, 89-109.
- SPIEGELBERG, Herbert "Towards a Phenomenology of Experience," APQ, vol. 1 (1964), pp. 325-332.
- STADLER, Ingrid H. "On 'Seeing As,'" PR, vol. 67 (1958), pp. 91-94.
- STEVENS, S. S. "The Direct Estimation of Sensory Magnitudes—Loudness," *American Journal of Psychology*, vol. 69 (1956), pp. 1-25.
- STEVINS, S. S. "Quantifying Sensory Experience" in P. Feyerabend and G. Maxwell (eds.), *Mind, Matter, and Method: Essays in Philosophy and Science in Honor of Herbert Feigl* (Minneapolis, 1966).
- STRANG, Colin "The Perception of Heat," PAS, vol. 61 (1961), pp. 239-252.
- STRAUS, Erwin "The Forms of Spatiality" in *Phenomenological Psychology* (New York, 1966).
- "Phenomenology of Hallucinations," *ibid*.
- STRAWSON, P. F. "Professor Ayer's 'The Problem of Knowledge,'" P, vol. 32 (1957), pp. 302-314.
- "Perception and Identification," ASSP, vol. 35 (1961), pp. 97-120. Symposium with S. Hampshire.
- STREANS, Isabel "The Grounds of Knowledge," PPR, vol. 2 (1941), p. 359.
- SWARTZ, Robert J. "Color Concepts and Dispositions," S, vol. 17 (1967), pp. 202-222.
- TAYLOR, Charles and M. KULLMAN "The Pre-objective World," RM, vol. 12 (1958), pp. 108-132.
- THALBERG, Irving "Looks, Impressions, and Incorrigibility," PPR, vol. 25 (1964), pp. 365-374.
- THEOBALD, D. W. "Observation and Reality," M, vol. 76 (1967), pp. 198-207.
- THOMAS, L. E. "Looking," PQ, vol. 7 (1957), pp. 109-115.
- TILGHAM, B. R. "Aesthetic Perception and the Problem of the Aesthetic Object," M, vol. 75 (1966), pp. 351-367.
- TOLMAN, E. C. "Principles of Purposive Behavior" in S. Koch (ed.), *Psychology: A Study of Science*, vol. II (New York, 1959).
- TOMAS, Vincent "Aesthetic Vision," PR, vol. 68 (1959), pp. 52-67.
- TUCKER, John "The Television Theory of Perception," BJPS, vol. 9 (1958), pp. 51-57.
- UNGER, Peter "Experience and Factual Knowledge," JP, vol. 64 (1967), pp. 152-173.
- URMSON, J. O. "Recognition," PAS, vol. 56 (1956), pp. 259-280.
- USHENKO, A. "A Note on Russell and Naive Realism," JP, vol. 53 (1956), pp. 819-820.
- VEER, Garrett, L. VANDER "Austin on Perception," RM, vol. 17 (1963), pp. 557-567.
- VENDLER, Zeno "Verbs and Time," PR, vol. 66 (1957), pp. 143-160.
- VERNON, M. D. "Cognitive Inference in Perceptual Activity," *British Journal of Psychology*, vol. 48 (1957), pp. 35-47.
- "Perception, Attention, and Consciousness," *Advancement of Science*, 1960, vol. III. Reprinted in P. Bakan (ed.), *Attention* (New Jersey, 1966).
- VESSEY, G. N. A. "Seeing and Seeing As," PAS, vol. 56 (1956), pp. 109-124. Reprinted in Swartz.
- "Unconscious Perception," ASSP, vol. 34 (1960), pp. 67-78. Symposium with J. P. Day.
- "Berkeley and Sensations of Heat," PR, vol. 69 (1960), p. 210.
- "Berkeley and the Man Born Blind," PAS, vol. 61 (1961), pp. 189-206.
- "Margolis on the Location of Bodily Sensation," A, vol. 27 (1967), pp. 174-176.
- WALLRAFF, Charles F. "Sense-Datum Theory and Observational Fact: Some Contributions of Psychology to Epistemology," JP, vol. 55 (1958), pp. 20-32.
- WALTON, Kendall "The Dispensibility of Perceptual Inferences," M, vol. 72 (1963), pp. 357-368.
- WATLING, John "About A. J. Ayer's 'The Problem of Knowledge,'" RIP, vol. 12 (1958), pp. 75-85.
- "Phenomenalism Flawed," I, vol. 6 (1963), pp. 196-199.
- WATSON, A. "Consciousness and Perception in Psychology," ASSP, vol. 40 (1966), pp. 85-100.
- WHEATLEY, John "Like," PAS, vol. 62 (1962), pp. 99-116.
- WHITE, Alan "The Causal Theory of Perception," ASSP, vol. 35 (1961), pp. 153-168. Symposium with H. P. Grice, reprinted in Warnock.
- "Attending and Noticing," PAS, vol. 63 (1963), pp. 103-126.
- "The Alleged Ambiguity of 'See,'" A, vol. 24 (1963), pp. 1-5.
- WHITELY, C. H. "Physical Objects," P, vol. 34 (1959), pp. 142-149.
- WHITMORE, Charles E. "Perception and Experiment," JP, vol. 54 (1957), pp. 401-409.
- WHITROW, G. J. "Why Physical Space Has Three Dimensions," BJPS, vol. 6 (1955), pp. 13-31.
- WHYTE, Lancelot Law "Some Thought on Certainty in Physical Science," BJPS, vol. 14 (1964), pp. 32-38.

335081

- WILLIAMS, C. J. F. "Form and Sensation," ASSP, vol. 3 (1965), pp. 189-154. Symposium with R. J. Hirst.
- WILLIS, Richard "The Phenomenalist Theory of the World," M, vol. 66 (1957), pp. 210-221.
- WITHERS, R. J. F. "Epistemology and Scientific Strategy," BJPS, vol. 10 (1959), pp. 89-101.
- WOHLWILL, J. F. "Developmental Studies of Perception," *Psychological Bulletin*, vol. 57 (1960), pp. 249-288.
- WOLFF, Robert Paul "Hume's Theory of Mental Activity," PR, vol. 69 (1960), pp. 289-310.
- WOLGAST, Elizabeth H. "Perceiving and Impressions," PR, vol. 67 (1958), pp. 226-236.
- "The Experience in Perception," PR, vol. 69 (1960), pp. 165-182.
- "Qualities and Illusions," M, vol. 71 (1962), pp. 458-473.
- "A Question About Colours," PR, vol. 71 (1962), pp. 328-339.
- WOLTERSTORFF, Nicholas "Qualities," PR, vol. 69 (1960), pp. 183-200.
- YOLTON, John W. "Philosophical Realism and Psychological Data," PPR, vol. 19 (1958), pp. 486-501.
- "Broad's View on the Nature and Existence of External Objects" in P. A. Schlipp (ed.), *The Philosophy of C. D. Broad* (Tudor, 1959).
- "Seeming and Being," PQ, vol. 11 (1961), pp. 114-122.
- YOST, R. M. "Price on Appearing and Appearances," JP, vol. 60 (1963), pp. 328-333.
- ZENER, Karl "The Significance of Experience of the Individual for the Science of Psychology" in H. Feigl, M. Scriven and G. Maxwell (eds.), *Minnesota Studies in the Philosophy of Science*, vol. II (Minneapolis, 1958).
- ZERNER, Karl and Mercedes GAFFRON "Perceptual Experience: An Analysis of its Relations to the External World through Internal Processings" in S. Koch (ed.), *Psychology: A Study of Science*, vol. IV (New York, 1962).
- ZIEDINS, R. "Conditions of Observation and States of Observers," PR, vol. 65 (1956), pp. 299-323.
- "The Possibility of Scepticism about Perception," PQ, vol. 16 (1966), pp. 329-340.
- "Knowledge, Belief and Perceptual Experience," AJP, vol. 44 (1966), pp. 70-88.
- ZINKERNAGEL, Peter "On the General Problem of Objective Relativity," M, vol. 72 (1962), pp. 38-45.

II. TRAITS OF CHARACTER: A CONCEPTUAL ANALYSIS

RICHARD B. BRANDT

RECENT philosophical psychology has paid scant attention to the concept of traits, either of personality in general or more specifically of character, in comparison with that devoted to concepts of choice, desire, will, and intention. This is unfortunate, for several reasons. Trait-names play a large role in the practical discourse of the ordinary person, whether in back-fence commentary or in letters of recommendation, and there is always a danger that things may go wrong if the speaker is totally vague about the analysis of the term. The same danger is run by historians who make free use of trait-concepts in explanations of events (the collapse of a nation may be ascribed to things like the queen's vanity). Furthermore, there is currently a large literature on traits by psychologists (the work, e.g., of writers like Cattell, Eysenck, McClelland, Allport, Heider, and Norman). Despite the mathematical sophistication of these writers, particularly the factor analysts, it seems that a little philosophical spadework at the foundations can serve at least to suggest interesting questions relevant to the large theoretical edifices that have been raised. Finally, and most important, the moral philosopher has a considerable stake in the understanding of character-trait-concepts. Some philosophers, like Ross and Hartmann, have included character-traits among the things that have intrinsic value. Other writers, concerned with the issue of determinism, have thought that the determinist thesis about human conduct arises from a confusion about the nature of traits of character. Last, there is the thesis that an act is morally blameworthy only if it would not have occurred but for some defective trait in the moral character of the agent, and that it is morally praiseworthy only if it would not have occurred but for some superior trait in the moral character of the agent—a view that is a development of Aristotle's suggestion (*Nicomachean Ethics*, Bk. II) that a necessary condition of an act's being virtuous is that it "be based on a fixed and permanent

quality" in the agent's character. Obviously this view makes no sense if the concept of a trait of character, or of degrees of such traits, makes no clear sense—or if it cannot be made clear how it can be, and be known, that a certain action would not have occurred but for the presence in the agent of some trait to a defective or superior degree. So, concern about the analysis of character-trait-concepts can hardly be written off as aimless devotion to lexicography.

In what follows I shall attempt to delineate the common logical structure of a class of terms I shall call "character-trait-names." More exactly, I shall offer a *schema*, as definite as possible, for the definition of all character-trait-names such that, by filling in the blanks appropriately, one would get illuminating explications. This statement of a program obviously needs explanation.

First, I assume that there are terms which it would be agreed are character-trait-names, including such terms as conscientiousness, considerateness, courage, generosity, honesty, kindness, modesty, prudence, reliability, responsibility, self-control, sympathy (compassion), truthfulness, and unselfishness. In contrast, there are terms designating traits of personality which I should not classify as names of traits of character, at least not as traits of *moral* character, such as: adventurousness, calmness, credulity, emotional instability, energy, fussiness, good nature, gregariousness, imagination, inflexibility, intelligence, optimism, pedantry, poise, polish, shyness, talkativeness, tenseness, timidity, and warmth. The schema to be offered may fit some of the latter group as well as the former group; what is here attempted, however, is simply a schema that fits all, or at least most, of the former group. Obviously it does *not* suit some of the latter group, e.g., "energetic."

It is important to notice that the schema is not in any sense intended to provide an analysis of the term "trait of character"; what is intended is a schema which will enable illuminating analyses

of various terms which are in fact names of traits of character.

It may be objected that the suggested dichotomy of personality-trait-terms into those designating traits of moral character and others is questionable, and there is no reasonably definite intuitively acceptable class of names of traits of moral character. To this complaint several concessions should be made. First, possibly there is a use of "character" very close to that of "nature," so that almost any feature of a whole person can count as a trait of character in that sense. Moreover, when we say a person is "quite a character" we mean that he has some distinctive eccentric features of some sort, and on the other hand when we say a person is a "man of character" we mean that he is a person of good character. But there surely is a use of "trait of moral character" which applies to a fairly definite set of traits, roughly identical with the set of what have traditionally been called moral virtues or moral vices, and in which "moral character" is used in a sense different from any of the foregoing. It is this sense of the term "character" (= "moral character") which philosophers have had in mind when they have accused Aristotle and Hume of discussing, in their accounts of virtues, some traits which are not traits of moral character at all, e.g., being witty. This sense is a familiar one; printed forms for testimonials often contain a space for comments on "character," in which evidently something else is to be described than intelligence, flexibility, emotional stability, etc., to which topics other spaces are devoted. Traits of character in this sense appear to have two features in common: (1) Each is a trait of personality which is normally, in any adult (not just in a person with a certain role such as a mother or a nurse or a lawyer), either an important asset or an important liability for cooperative living, from the point of view of society. (2) An expression of any of them in action is within voluntary control, in the sense that a person always can produce it given appropriate interests (wants, aversions) on his part.¹ Accordingly, some trait which might be counted a virtue in a military commander or a burglar, say daring, is not usually classified as a trait of character, certainly in the sense of "moral character." It cannot be denied

that the phrase "trait of character" carries rather specific implications for some people. Some would not count the natural sympathy of a six-year old, shown on the playground, as a trait of character; they would say that a trait of character must be learned through practice based on moral reflection.

We may have to accept the fact that there are terms (perhaps "patience") about which our linguistic intuitions are mute, about which it is not clear whether they do or do not have the features characteristic of traits of moral character. If so, our success in attaining our present objective is not threatened, for the chief goal of the analysis is illumination of paradigm, well-recognized examples. For the same reason we need not be perturbed if, as I think is the case, evaluative judgments are required for some decisions about what is to count as a trait of character.

But in what sense is it proposed that our schema will provide an "illuminating explication" of these terms? It will not be asserted that all the features of our schema are required by the "conscious meaning" of these terms in the way in which the definition in terms of "unmarried male" is required by the conscious meaning of "bachelor." It is claimed that the schema is consistent with such conscious meanings, and at least roughly consistent with intuitions about the application of the terms by language-users in concrete cases. More important, however, the schema is intended to be consistent with the conceptual framework employed by both common sense and scientific psychology; it is intended that the terms, as defined, are ones which we shall be happy to continue using, given our total psychological theory. As such, our proposed definitions may expand, make precise, or possibly even somewhat change our present intuitive use of the terms, hopefully in a useful way.²

I. CHARACTER TRAITS AS DISPOSITIONS; REJECTION OF THE "SUMMARY" VIEW

It is natural to regard trait-names as naming dispositions of a person. Just as we might explain "x is soluble in water" as meaning "if x were placed in water it would dissolve," so it is natural to

¹ See the writer's *Ethical Theory* (Englewood Cliffs, N.J., Prentice-Hall, Inc., 1959), pp. 466-468. The proposal is somewhat similar to that of P. H. Nowell-Smith, *Ethics* (Baltimore, Md., Penguin Books, 1954), pp. 300-306.

² Some writers seem to think that the main content of trait-names is just the evaluative, or praise-blame element. This has been shown to be mistaken by Dean Peabody, "Trait Inferences: Evaluative and Descriptive Aspects," *Journal of Personality and Social Psychology*, vol. 7 (1967), pp. 1-18.

propose—roughly following one of Ryle's suggestions—that "*x* is vain" amounts to a set of subjunctive conditionals, one of which might be, "if it occurred to *x* that doing *A* would likely secure the admiration and envy of others, he would be strongly tempted to do *A*."³ Such a formulation implies that trait-attributions support counterfactuals: even though *x* were not in fact to find what he thinks is a device for arousing the admiration of others, he *would* be strongly tempted to utilize it, if he did.

A dispositional view of trait-names will here be defended, but with the reservation that a more precise account would construe them as theoretical terms. Exactly how this would be will be indicated later. But for the purpose of getting clear the reasons for preferring one type (what I shall call a "motivational dispositional" analysis) of analysis to some other quite different types, extended pursuit of this rather subtle refinement is unnecessary and would only be confusingly complicating.

It should be noticed that a dispositional analysis along the above lines does not exclude the use of probability or frequency notions in the analysis. One might, for instance, propose an analysis such as "if it occurred to *x*, . . . , then he *probably* (or "very frequently" or "relatively frequently") would. . . ." How the term "probably" (etc.) should be construed naturally would need explanation, by an advocate of such a proposal. In fact, however, one of the things I shall be doing is offering reasons for rejecting probability/frequency ideas from the analysis.

A dispositional analysis of some form would, I believe, be acceptable to most psychologists interested in the theory of traits of personality. Some philosophers, however, are disposed to adopt a form of what I shall call the Summary Theory. The first thing I wish to do is to give reasons for dropping the Summary Theory.

The Summary Theory may take either of two forms; *pure*, or *mixed*. The *pure* form holds that to ascribe a trait to a person is simply to affirm that a certain corresponding form of behavior/experience has occurred, in the person, frequently or relatively frequently in the past—perhaps with the "implication" in some sense that the same frequency may be expected to continue. Thus, to say

that a person is witty is to say that he has relatively frequently succeeded in making amusing sallies in the past—and perhaps to imply that this frequency may be expected to continue.⁴ The *mixed* form construes trait-names partly as dispositional, thus adopting some subjunctive conditional expression as *part* of the analysis, but insists that to this we must *add* some expression to the effect that in fact manifestations of the trait have occurred in the individual's past, either frequently or relatively frequently.

Three main types of reason appear to be offered in support of some version of the Summary Theory. First, it is said that "normally" we make trait-ascriptions only when we know about actual manifestations of the trait. And this seems true; but it is hardly an objection to a dispositional theory. For it is obvious that *normally*, viz., in the absence of test information like the Rorschach, or the MMPI, etc., we do and must base our trait-ascriptions on observations of relevant behavior manifesting the trait. But this would be true even if the dispositional analysis were correct, and it can hardly be an objection to the dispositional analysis. Secondly, it is argued that we can be *certain* of a trait-ascription only when we are able to cite numerous instances of past behavior, and that such information, and only such, has *conclusive* force in appraising a trait-ascription. In the case of this argument, we can concede that *other* evidence may not be conclusive or lead to certainty; psychological tests, for instance, are not that reliable. Even if a psychologist had a battery of test results in front of him—Rorschach, Thematic Apperception, MMPI, etc.,—and also biographical data about traumatic experiences in the person's childhood, rejection by his parents, etc., and even if all this evidence pointed firmly to the likelihood that the individual was an anxious person, it is still *possible* that he would respond to life-situations with less anxiety than the normal person. Highly unlikely, but possible. But there is no reason why the dispositional theorist need deny this; he can admit that past real-life reactions are a more certain guide to present dispositions than any psychological test, but go right on affirming that what trait-ascriptions do is affirm present dispositions and not facts about the past. (Inciden-

³ See Gilbert Ryle, *The Concept of Mind* (London, Hutchinson's University Library, 1949), p. 89. Ryle offers somewhat different formulations.

⁴ Something at least like this view has been defended by Stuart Hampshire, "Dispositions," *Analysis*, vol. 14 (1953), pp. 5-11; George Pitcher, "Necessitarianism," *The Philosophical Quarterly*, vol. 11 (1961), esp. pp. 207-208; and Betty Powell, "Uncharacteristic Actions," *Mind*, vol. 68 (1959), pp. 492-509, esp. pp. 500, 502.

tally, past real-life reactions are not necessarily conclusive, or certain, evidence about traits. For instance, a really anxious person may meet a situation, which would even be anxiety-arousing to a normal person, after some elating experience such as an unusual achievement on his part, and he may *not* react to it with anxiety. It could well be, for such reasons, that a mass of psychological tests would be better evidence for a given person's traits than data about his actual past behavior in real situations. Thirdly, and finally, there is an argument which is simply an appeal to meanings: it would be contradictory to say that a person is *T* (has a certain trait) but has never behaved in a *T*-like manner. As W. P. Alston has put it, "A person who has never obeyed any orders might be correctly called 'potentially obedient,' 'an obedient type,' or 'a person who would be obedient if he had the chance'; but we could not be justified in terming him 'obedient' *tout court*. The occurrence of some instances of the correlated manifestation category is a necessary . . . condition for the application of the trait term."⁵

Is it *contradictory* to affirm that a person is *T*, or, on the evidence probably *T*, and at the same time to say that certainly or probably he has never acted in a *T*-like way in the past? I fail to see that it is, at least for the traits of moral character with which we are concerned. Such questions are, of course, difficult. But take "courageous." Suppose we knew a given person had lived a very sheltered life and had never been required to act in the face of a serious threat. (It is not easy to imagine comparable situations in the case of most traits; for the normal life-situations of human individuals are such that, in the case of most traits, if a person has the trait he cannot have failed to manifest it in behavior.) Would we infer of such a person that he *cannot* be courageous? Surely not. Indeed, there are conceivable psychological tests such that, given a certain result on these tests, we would say that the person is *probably* a courageous person. (Suppose there were a known high correlation between this trait and the presence of a certain chemical element in the blood, and tests showed this chemical to be present in the blood in ample quantity.) It is true that, as things now stand, we would hesitate to affirm roundly, without any "probability" reservation, that a person was

courageous without any behavioral evidence; but this fact shows something, not about the meaning of "courageous" (or other trait-names of interest to us), but about our convictions on what is adequate evidence for trait-ascriptions; and if that is the point, then this argument reduces to the second argument. There is no evidence to suggest that it is self-contradictory to say that a person is *T* but has not behaved in a *T*-like way.

The objections to a dispositional theory, then, seem at best highly dubious. There is also, however, an argument which appears to show that the Summary Theory is simply wrong. This argument adduces as its premiss the fact that we often draw inferences about a person's traits of character, on the basis of a *single* piece of behavior.⁶ For instance, if a young boy is threatened with a beating by larger boys if he fails to do a certain thing, and he steadfastly refuses, we justifiably assert that he is courageous. It is true that we do need a good deal of information to eliminate other hypotheses—such as that he mistakenly thought he had nothing to fear, or that he had just taken a courage-drug, etc. But how could we draw such an inference, with high confidence, from any amount of information about a single situation if trait-affirmations were assertions about the frequency of behavior in the past? (The present behavior is, of course, one case; but to say that a person is courageous is surely not to say merely that he has acted courageously once.) It is true that we could draw such inferences on such evidence *if* we were assured of some general proposition to the effect that people do not behave in *this* way unless they have frequently behaved in a comparable way in the past. But we do not have evidence that such generalizations are true when we are drawing such inferences; and it is doubtful whether they are true. (It seems there could be a first time when a person manifested his courage.) Surely we do not appeal to any such generalization in drawing or justifying our inference from the behavior. We make the judgment of courage because that hypothesis is the most plausible explanation in view of our total information about the situation and about how people generally behave; and we justify our judgment as being such. The mode of our inference, and of our justification of our inference, simply does not jibe with the Summary Theory.

⁵ W. P. Alston, *Toward a Logical Geography of Personality: Traits and Deeper Lying Personality Characteristics* (forthcoming).

⁶ This point was noticed by Maurice Mandelbaum, *The Phenomenology of Moral Experience* (New York, The Free Press, 1955), p. 147.

II. TRAITS OF CHARACTER ARE RELATIVELY PERMANENT DISPOSITIONS

Traits of character are relatively permanent features of a person. So, if we are to give a dispositional account of character trait-names, we must begin with some such phrase as "It is a relatively stable and permanent feature of *X* that, were he . . . , he would" There is none of the trait-names of interest to us which we would apply to a person in virtue of some relatively unstable and temporary feature. Furthermore, traits of character do not undergo cyclical modifications like the "needs" for food or water or sex; it is not as if, having acted sympathetically (assuming being sympathetic is a trait) at noon, I shall have a desire for more sympathetic action at six o'clock and not before. Traits do, of course, change; we often have occasion to say that when a boy Mr. *X* was so-and-so, but as an adult he has become such-and-such, where the latter is incompatible with the former.

III. TRAITS OF CHARACTER ARE INTRINSIC WANTS/AVERSIONS

I now wish to claim that traits are relatively permanent dispositions of a *specific kind*—the kind of dispositions that wants and aversions are. We shall see that there is more to be said; for instance, a given trait-name implies that the relevant desire/aversion reaches a certain level of intensity. These refinements will be discussed later. In the present section I shall discuss the central claim that trait-names designate desires or aversions (or some structure of these), and indeed ones that are *intrinsic* in a sense to be explained.

This claim is certainly controversial; it would be contradicted not only by many philosophers but by many (by no means all) psychologists interested in traits. That it is a defensible claim is a central thesis of this paper. I shall call the claim "the motivational theory" of character-traits.

The motivation theory of character-traits, in claiming that traits are dispositions of the want/aversion kind, is assuming that wants/aversions are themselves dispositions (subject to a refinement already noted, which will be discussed further below). I do not know that this assumption needs any apology.

Let me begin by pointing out that there is some *prima facie* reason for thinking that a motivational theory of character-traits must be correct. Traits of character, as contrasted with "stylistic" traits like being affected, analytical, cheerful, clumsy, energetic, enthusiastic, excitable, or expansive, are concerned with *intentional actions*. Now, if we assume that character-traits are wants/aversions, we shall be in a position to do what we think we can properly do—*explain* intentional behavior by reference to character-traits. Notice that we do sometimes, or often, explain behavior (which is unusual enough to call for an explanation) by appeal to such things as that the person is unusually conscientious, or sympathetic, or considerate. So, if we adopt the motivational theory, we can properly explain intentional behavior in the way we do explain it, at least in thoughtful moments. That is some reason for adopting the motivational theory. Of course, it may be that if we construe character traits in some different way, it will still be possible to appeal to them as in some sense explanatory of intentional behavior. This is one thing we shall be looking into.

If we look at the literature on the theory of action and motivation,⁷ we find a list of variables, of which intentional action is thought to be a function; the function it is of these variables is the law or laws of behavior. Now if we assume that traits of character figure in the explanation of actions, and if we look among the psychologists' list of factors for items which might represent specific traits of character, the most plausible candidate is surely what is sometimes called "need" or "drive" or "want," of which the "need" for food or affiliation or achievement or security or freedom from pain (aversion to pain) are prime examples. If traits are construed as specific kinds of need (want/aversion), at least an understanding of the role they play in action would be a straightforward matter; we could see how they fit into a widely accepted pattern of psychological explanation of intentional action.

What kind of "pattern of explanation of intentional action"? It must be admitted that motivation theorists talk somewhat different languages, e.g., the behavioristic learning theorist emphasizing "anticipatory goal responses." We can ignore these differences, however, if we assume, as seems plausible, that these languages are intertransla-

⁷ For instance, J. W. Atkinson, *An Introduction to Motivation* (Princeton, D. van Nostrand and Co., 1964); or E. C. Tolman, "A Psychological Model" in Talcott Parsons and E. A. Shils (eds.), *Toward a General Theory of Action* (Cambridge, Harvard University Press, 1954).

table. The "pattern of explanation," put in one kind of terminology, consists in acceptance, as a general law of behavior, of the thesis that (roughly) a person always acts (moves his body) in such a way as to *maximize expectable utility for himself*; and in assimilating all actions, given the context of beliefs/needs of the agent, to this law as instances of it. To spell out a bit what is meant, we construe "expectable" in the following way: The organism believes itself to be in a certain location and situation, and it believes that there are various states of affairs which could be produced by various basic actions (movements) it could produce, with varying degrees of probability. Or, in other words, the organism associates various outcomes as connected with various of its possible actions, with differing degrees of firmness or confidence. We construe "utility" as follows: Various needs (wants/aversions) of the agent are *in force* (see below) at the time of action, with different degrees of strength. These needs make "contact" with various ones of the conceived possible outcomes, which are states of affairs toward which the organism has wants/aversions; as a result there is generated a psychological "valence" of that state of affairs, positive or negative, with a strength corresponding to the degree of strength of the respective need at the time. Different outcomes which make contact with the same need, however, somehow have different incentive values (e.g., a choice steak having more incentive value in relation to the hunger need than does goulash), and these differences are also reflected in the size of the "valence." (There are difficulties, definitional and otherwise, in this concept of lumping need-strength and incentive value together in the "valence.") Now we think of assigning numbers to the valence of an outcome and to the subjective probability of that outcome, given a certain action; and the product of these numbers represents the "expectable utility" of the action thought to lead to the outcome. (The same action might lead to several outcomes; the expectable utilities then have to be summed.) The size of the expectable utility may be represented as a force-vector in the direction of a given action. The "law" of behavior says, roughly, that the organism will take that action which conforms to the strongest force-vector. If we employ this kind of law as our pattern for explanation, then we shall, for instance, explain my walking to the refrigerator by such things as how thirsty I am,

what kind of drink I think is available in the refrigerator (beer, warm water), how anxious I am not to be interrupted in writing a given paragraph or finishing some job, how likely I think it is that if I walk toward the refrigerator I shall be able to open it, etc. Similarly for traits of character. If, for instance, we construe sympathy as an aversion to other people being in distress, we shall fit it into the pattern of explanation much as we do thirst; what I shall do if I see a child fall off his bicycle will be a function of the strength of this aversion, how serious distress I take it the child is in, how little I want to be interrupted (maybe I am in the midst of a crucial game of tennis), and how likely I think it is that I shall be able to relieve the distress if I move to the rescue; and so on.

The foregoing is only a simplified sketch of the relation between the intentional basic movements of a person and the values of certain variables characterizing him at the time. Even as such, a few further points should be noted about it. For one thing, "expectable" has to be defined as "subjectively probable" and explained in such a way that what is expectable for a person can be ascertained, at least partially, independently of what he does at the moment, else the "law" is analytic. Furthermore, a certain need (want or aversion) may characterize a person at a time but, because he has forgotten about it, it may fail to influence his behavior. It seems necessary, for this and other reasons, to distinguish between "latent" and "active" needs (ones *in force*) of a person at a time; and the needs in force, again, must be determinable by reference to other phenomena than the person's consequent behavior (or, at least, partially by reference to other phenomena), again in order to avoid rendering the "law" of behavior analytic. Again, it appears that motivation theory somewhere has to allow a place for such variables as "impulsiveness," which are not identifiable either with needs or beliefs. Moreover, it has to be admitted that psychologists are much more at home in behavior-theory applicable to non-symbol-using animals than for men, and do not have much to say about a process of reflection issuing in behavior. In general it seems necessary to distinguish between a simpler model for the explanation of behavior and a more complex one in which symbolic processes play a significant role, as was argued by the writer and Jaegwon Kim in an earlier publication.⁸

⁸ R. B. Brandt and Jaegwon Kim, "Wants as Explanations of Actions," *The Journal of Philosophy*, vol. 60 (1963), pp. 425-435.

Fortunately it is not necessary for us here to be clear about the nature, or even the general form, of an acceptable theory of human action. For our purpose is simply to get clear what is intended by saying that a character-trait-name designates some need (want/aversion) or, more exactly, by saying that ascription of a character-trait functions to assert something about the needs, or level of intensity of needs, of a person; and to get clear in what general way the ascription of character-traits (or, more generally, needs) can function in the explanation of behavior.

It may be that we must acknowledge various differences between the needs designated by character-trait-names and those most familiar in behavior theory. Compare, for instance, the need thirst and the trait sympathetic. First, intensity of thirst seems to be correlated with the dehydration of the body cells, etc., hence a function largely of hours since water-intake, exercise, etc., whereas sympathy is not a function of such variables. Again, thirst decreases in strength rather rapidly as the consummatory activity goes on; after a couple of glasses of water the thirst of a human being in normal circumstances drops to zero and resistance to further drinking develops. This satiation phenomenon is evidently not so marked, if it exists at all, with the sympathy need; if a second child falls off his bicycle five minutes after the first one, one does not find it overpoweringly repulsive to go to his rescue also. (Needs like affiliation and achievement are less variable than thirst, although satiation is clearly observable in their case too.) In this respect sympathy rather resembles the aversion to pain—since it appears that a continuous dose of pain does not decrease one's interest in avoiding it. In general it seems more plausible to construe character-traits as aversive needs than as positive wants.

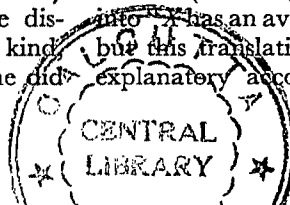
My proposal has been (as a first approximation) that character-trait-names designate dispositions; and it follows from this that (at least as a first approximation) they can be explained by appropriate subjunctive-conditional statements. Since I am also suggesting that the disposition in question is a kind of need, it follows that the subjunctive conditionals by means of which trait-names can be explained will be of the type by which need-concepts can be explained. So, *if* to say that a person *wants* something of a certain sort is to say that under certain conditions he would be disappointed (if he didn't get a thing of the kind, when he thought he would) or feel joy (if he did

get a thing of the kind when he thought he wouldn't) or tend to act in a certain way (which he thought would bring him a sample of the situation wanted and would not prevent other things he wanted more), etc., then to say that a person has a certain trait of character will be to make a statement that can be explained along similar lines.

There is, however, a complication; and it is time to state more explicitly what would be claimed if we were not limiting ourselves to a "first approximation." In an earlier paper (see footnote 8), the writer and Jaegwon Kim defended the view that "want" and "aversion" are not, strictly speaking, disposition terms but theoretical terms, and hence that in principle no producible set of statements relating them to experiences like disappointment will provide an explicit definition of them. It was argued that many important single statements of this sort are neither analytic nor synthetic. Moreover, it seems impossible to spell out exactly the conditions under which a given want/aversion will lead to disappointment, etc., certainly if there is permitted no reference to *other* wants/aversions, or to *beliefs*—a concept with a status like that of "need." These strictures presumably hold for character-trait-names. Hence my suggestion is that for them, as for need-statements generally, true statements relating them to experience have to be regarded as correspondence rules, or dictionary statements, in the sense in which these terms are used of theoretical language, rather than as explicit definitions. These statements cannot be construed as giving an acceptable explicit definition; on the other hand, they are not simply synthetic, since the meaning of the theoretical terms is determined by their role in such true statements.

Having said this, however, I propose from now on mostly to ignore it, and to write as if character-trait-names name dispositions which can be explained by means of counterfactuals. It should not be forgotten, however, that when such things are said, it is intended that the statement be viewed only as a first approximation.

It may be helpful, as an illustration of the present conception of traits, to indicate how an expression like "*X* is sympathetic" might be partially explained by a set of subjunctive conditionals. (One might translate "*X* is sympathetic" into "*X* has an aversion to people being in distress"; but this translation does not provide the kind of explanatory account we are looking for.) Sub-



junctive conditionals which we might regard as partial substitutes for "is sympathetic" would presumably include the following: (1) "... would feel disturbed, other things being equal, if he perceived some sentient being to be in acute distress." (2) "... would feel relieved, if he perceived a being in distress in process of being helped, provided he had earlier felt discomfort at the person's distress." (3) "... would be motivated to relieve the distress, if he believed that he could do so and that no one else would if he did not."⁹ (4) "... would feel guilty, other things being equal, if he perceived distress he thought he could relieve but did not, provided justifying or excusing considerations were absent." (5) "... would notice a case of distress, if he were presented with it perceptually." (6) "... would remember a previous case of distress, if he had noticed it before, and were now in a position to give relief." Doubtless the reader will be able to think of various qualifications that must be added, before he would regard any of these statements as true of a person who is sympathetic.

Expressions like the above cannot be used to explain all character-trait-names, for some of them, e.g., "dishonesty," refer to the lack of a trait which might be explained as above. Dishonesty is a trait name, but it designates the absence of honesty, so that if we are to try to explain it in terms of statements somewhat like the above, they have to take the form, "It is not the case that if *X* . . . , then *X* would" Moreover, just as this vice has to be viewed as the absence of a virtue, so some virtues have to be viewed as the absence of a vice, e.g., courage as the absence of cowardice (as will be explained below). There may have to be more complications still. All that is here claimed is that character-trait-names function as ascriptions of some state of the person's *system of wants/aversions*, and can be explained accordingly. Only in this sense is our theory "motivational."

One perplexing question which may be raised about the motivational proposal is appropriately mentioned at this juncture. Why is it that the analysis of most character-trait-names involves the notion of guilt feelings? Guilt feelings do not appear in the analysis of most personal wants and

aversions, e.g., wanting a drink or wanting a raise in salary, or aversion to bananas. In the case of personal wants/aversions, the notions of disappointment and anxiety may well occur in the analysis, but not guilt feelings. Why is there this disparity? I raise this question, without knowing the answer to it.

In the title of the present section, it was indicated that the wants/aversions which are identical with traits of character must be "intrinsic." What was meant by this is that the truth of the "if . . . then . . ." statements which constitute the explanation of the trait-name is not *derivative*, or at least not *wholly* derivative, from some other want/aversion. For instance, a person might be motivated to give to charity, and as such we might be inclined to call him generous. But if his willingness to give to charity were dependent on his desire to be in the public eye, or to improve his reputation, or to be elected to public office, we should not call him generous. (The term "derivative" might here be defined in various ways; I propose to ignore problems of its definition.)

So much for the statement of the proposal. It should be noticed that the proposal has an implication for psychology of some interest, which might permit empirical testing of it: for it appears that wants/aversions, when they are learned, are learned in ways different from behavioral responses. So, if traits of character are needs (wants/aversions), the development of them will follow the same laws as the general laws for learning needs.¹⁰

The foregoing motivational construction of character-trait-names is not self-evidently true, and I shall now produce some support for it. But first, what is the alternative, assuming we have dismissed the Summary Theory? Consideration of this will bring out the distinctiveness of the theory here proposed.

Of course, strictly, there is not just *one* alternative. One could even hold—indeed it has been held—that character-trait-names are primarily evaluative, and that the cognitive content is minimal, even negligible. Again, it might be said that the proper lines have not been drawn with respect to the definition-theoretical-term distinction. I think,

⁹ This explanation is circular, if, as appears to be the case, "is motivated" has to be taken to mean, in part, "would take action . . . if he did not want more strongly to do something else." In the present context, this circularity is not an objection.

¹⁰ See, for instance, E. C. Tolman, "There Is More than One Kind of Learning," *Psychological Review*, vol. 56 (1949), pp. 144-155; R. B. Cattell, *The Scientific Analysis of Personality* (Chicago, Aldine Publishing Co., 1965), ch. 10; D. C. McClelland, *et al*, *The Achievement Motive* (New York, Appleton-Century-Crofts, 1953), pp. 89-96; O. H. Mowrer, *Learning Theory and Personality Dynamics* (New York, Ronald Press, 1950), *passim*.

however, that these possibilities are unimportant, and I shall ignore them. Where will important differences lie? Presumably there can be differences of opinion about specific proposals I shall make below about specific traits; but if the disagreement is about a pattern of analysis, it must be about some broader feature common to the analysis of the whole set of traits. Obviously *any* theory of traits of character is going to construe them as especially related to, and in some way explanatory of, intentional behavior (among other things); so the contrast must be between the motivational theory and one or more other theories with this feature in common. What is characteristic of the motivation theory is that it construes traits of character as like (as playing a role in common-sense and scientific psychological theory alike), say, the need to achieve, or the aversion to the disapprobation (or disapprobative behavior) of other persons. As such, the motivation theory relates traits to intentional behavior rather indirectly; it holds that, under conditions not fully understood, they become active and generate "force vectors" in the psychological field of the person, their direction and degree depending partly on the person's beliefs; what the person actually does is a function of the force vectors in his psychological field at the moment of action. Furthermore, as was pointed out by A. F. Shand many years ago, the existence of needs/aversions brings with it relevant changes in many of the prospective/retrospective emotions of a person, as his psychological field changes—such emotions as joy, anxiety, sorrow, despair, disappointment, and hope. There is, then, a large and fairly definite conceptual framework which traits must fit, if they are what the motivation theory says they are. The point about how motives are learned, mentioned just above, is a part of this. But, if we abjure motivation theory, *what* will be the connection between traits and behavior (or something else), which presumably will be definitive of that conception of traits? One thing is clear; such a theory will not make use of the explanatory framework of motivation theory; the conception of a psychological field in which there are behavior-forces which are a function of the whole set of needs active at the moment, and also of the agent's beliefs. In forming a conception of a more positive competing theory, it would be foolish to suggest that there may not be many, and possibly of quite different types. But the obvious

alternative—at any rate the one that seems actually to be espoused, as an alternative to the motivation theory—is a theory which proposes that *for various trait-names a form of behavior typical of that trait can be identified* (as talking for talkativeness), and that what it is for a person to have a certain trait is primarily for him to be *disposed to behave in the correlated typical way, in certain conditions, relatively frequently*. (Another way of putting it is to say that to have a certain trait is for the probability to be relatively high that the person will behave in the correlated way in certain conditions.) A trait can then be construed as a *set* of dispositions; not only as a disposition to act, but, under certain conditions, to have certain emotions relatively frequently—in this way absorbing whatever truth there is in the motivation theory hypothesis that when a person has a certain trait he will *feel* in certain ways in certain conditions. On this view, a trait is not a motive which may issue in a certain type of goal-seeking behavior depending on what other motives are present; it is a disposition for a relatively specific type of behavior to occur.¹¹

It should be noticed that this theory, which I shall call the Direct Disposition Theory, is a loosely defined theory, and it could be stretched so as scarcely to be distinguishable from the motivation theory. For instance, the "behavior" in question might be to do something from a certain motive, or for itself, or with some end in view. Or, more broadly, the phrase "in certain conditions" could be explained in such a way, in specific cases, as to specify all the beliefs and motives which must be present or absent for the behavior to (probably, or frequently) take place; in this case, the theory would be only terminologically different from the motivation theory. A type of Direct Disposition Theory, to be different in substance from the motivation theory, must propose explications of character-trait-names which do not in substance incorporate the conceptual framework of motivation theory; it will not, in substance, construe traits as playing the role of needs in motivation theory, or appearing in laws or general statements, or implying these, just as would be done if the traits were construed as needs/aversions.

On this alternative theory, it should be noted, explanation of behavior will take a different form, being more like the kind of explanation solubility in water gives of the soluble thing's actually dissolving in water. It is not suggested that such

¹¹ In this contrast between the two types of theory, I have benefited from Professor Alston's paper referred to above, and from numerous discussions with him, about this and related problems.

explanations are unimportant. Indeed, there can be and is a "science" of traits in this sense. For—but I am not suggesting there are not other possibilities—there is factor-analysis which can/discover interdependencies among traits and analyzes human personality in terms of a few statistical creations called "source traits."

I shall devote the remainder of this section to supporting the motivation theory. I shall proceed as follows. First, I shall consider two important objections to the theory, and meet them; I shall then show how one of the objections backfires and provides support for the motivation theory, and I shall then adduce another powerful reason for accepting the latter. Finally, I shall discuss some important character-traits which look initially like counterexamples and consider whether they may be construed compatibly with the motivation theory. We should remember, however, that I am not necessarily arguing that the motivational theory is true for *all* traits of character; it could be an illuminating and sound proposal for some but not all. Of course, if it is the account we must adopt for all, its interest is considerably increased.

I begin with two objections to the motivation theory. It is useful to cite a passage from R. S. Peters which incorporates the two objections in question, and at the same time appears to be a defense of a form of Direct Disposition theory.

We can explain a man's action in terms of traits of character, like considerateness, and punctuality. These may be reasons why people act; but they are not motives. For such terms do not indicate any definite sort of goals toward which a man's actions are directed. . . .¹²

Such terms express simply a correlation between typical situations and behaviour appropriate to them. To say that man acts because of indolence, vanity, or honesty is to relate his behaviour in a certain situation to a host of similar responses in similar situations—predominantly social ones. The man may be motivated in different ways while exhibiting the same trait; he may exhibit the same or different traits in bringing about a variety of aims. . . . The "trait" is a typical low-level concept useful for social intercourse, evaluative assessments of character, novelist's descriptions of character, and so on.¹³

Here there are two main arguments against the motivational theory: first, that trait-ascriptions obviously do not imply any particular goals (in

our terminology, "needs") at all; and second, that one may exhibit a given trait (e.g., act considerately) from very different motives so that having a given trait could not be identified with any particular want/aversion.

The first of these criticisms takes as being obvious the falsity of the motivation theory, except possibly as a "reforming" analysis of trait-names—the "reform" presumably being undertaken in the interest of construing trait-names so that traits can play a significant role in the explanation of behavior. But is it *obvious* that character-trait-names indicate no definite sort of goals? Surely we need to examine a representative list of such terms and consider whether they can plausibly be construed to refer to goals or wants/aversions. In the case of "considerateness," one would think that concern for discomforts, embarrassments, etc., of other persons precisely is implied by use of the term. Peters' other example in this context, punctuality, is not a trait of character at all. Peters' contention may seem plausible because most character-trait-terms name aversions rather than wants/desires, and hence it is true that in a sense they are *aiming* at nothing specific, as distinct from simply *avoiding* something specific, in *some* way or other. Still, aversions do have specifiable goals in an appropriate sense; for instance, while in a sense the aversion to pain has no goal, it is still true that there is a goal—there being *no* pain. Peters may have overlooked these facts, assuming that any need/want/motive must have a structure like the physiological needs, with a definite consummatory occasion after which the "tension" is diminished.

Peters' second criticism, that character-traits are dispositions to perform certain sorts of action and are not wants/aversions, since several actions may manifest one and the same trait but be motivated in diverse ways, is found also in M. H. Mandelbaum's work. Mandelbaum says:¹⁴

When we hold that a man is courageous we do not attribute any particular motive to him: his courage may be due to ambition, to emulation, or even to a fear of certain forms of opprobrium, and yet his action may be justly called courageous. . . . This trait [courage] may be grounded in such disparate conditions as combativeness, a calculated use of daring to attain certain personal ends, a craving for self-aggrandizement through social approbation, or a readiness to sacrifice one's self to an ideal.

¹² *The Concept of Motivation* (London, Routledge and Kegan Paul, 1958), p. 32; see also p. 5.

¹³ *Proceedings of the Aristotelian Society*, Supplementary Volume XXVI (1952), pp. 156–157.

¹⁴ *The Phenomenology of Moral Experience* (Glencoe, Ill., The Free Press 1955), pp. 142ff.

Mandelbaum lists fidelity, prudence, temperance, and sloth as further traits of this sort.

In reply to this objection I shall shortly suggest that some of Mandelbaum's examples can after all be construed straight-forwardly as aversions. But first let me point out why, in the case of *some* traits, a diversity of motivation is quite compatible with the motivation theory. This fact derives from the polarity (already adverted to) of some trait-pairs. For instance, suppose honesty is construed as aversion to certain things such as appropriation of the property of others, deceit, etc. But honesty has a polar opposite: *dishonesty*. But what is dishonesty? Presumably it is just the *absence* of the kind of aversion which constitutes honesty. As such, dishonesty is not identical with any particular want or aversion. Not, for instance, with acquisitiveness, which may be present in an honest man. Therefore behavior manifesting dishonesty *of course* will be variously motivated; the reason why it is called dishonest behavior is simply that it is behavior that would not have occurred had the aversion distinctive of the honest man been present. Any character-trait which consists in the *absence* of some kind of aversion typical of its polar opposite will necessarily be manifested by behavior motivated in all sorts of ways—by desires to do all sorts of things, which would not have controlled behavior in the presence of the aversion typical of the opposite trait. Among Mandelbaum's examples, both sloth and courage appear parallel with dishonesty in this respect; sloth being the absence of the aversion which constitutes ambitious industriousness, and courage being the absence of the aversion which is cowardice—the aversion to death, bodily injury, or damage to fundamental features of one's position. (I shall return to courage.) When we take this polar relationship into account, certain traits which look like counterexamples to the motivation theory turn out really not to be so.

I suggested above that the second objection backfires and points to something which is support for the motivation theory. What I mean is this. If the motivation theory is correct, then behavior manifesting the trait will be *any sort of behavior* which requires to be explained by reference to the need identical with the trait. So, if the motivational theory is correct, we may expect it to

be impossible to work out a thesis of the alternative theory: that some *definite kind of behavior* may be identified, correlated with a given trait. What one may expect is that no such definite correlated type of behavior can be identified, except by some reference to needs/aversions. What are the facts?

The Direct Disposition theory is at its best with traits like talkativeness (not, of course, a trait of character). For one may construe "*X* is talkative" as "*whenever X* has an audience, he is apt to talk a lot." In this case the behavior manifesting the trait seems easy enough to specify. Even here, however, some complications arise, for we do not regard a hotel clerk as talkative just because the conditional is true of him; we need at least some further clause like "in a normal situation."

When we consider some traits of character, however, the situation is more complex. Let us consider whether we can state what specific kind of behavior, identified without reference to motivation, may be expected to be relatively frequent, in specifiable conditions, in the case of a *generous* person. (The same would be true with virtually all of the traits listed in the third paragraph of this paper.) One might say that a generous person is one disposed,¹⁵ among other things, to make frequent gifts or perform services for others. But this cannot be right: for a generous person might not have contact with needy persons or causes, or might not have the means to make gifts or perform services. So at least we must introduce a condition: "*If* the person recognized a need, and believed he had requisite means at his disposal, he would, relatively frequently, make a gift or perform a service." But must one really do this *frequently*, to be generous? Suppose one saves one's resources over many years to make one large gift; this would seem to show generosity. It is at once clear that any one of a great variety of forms of behavior is enough: *any* form of behavior intended to reach some goal of a certain type. Again, a person with large means or capacities must do more to show himself generous than a person with limited means or capacities. (Example: the Widow's mite.) How much more? Apparently the answer is: Enough to prove the strength of one's motivation! Moreover, the making of gifts does not establish one's generosity unless they are made in a certain spirit, without ulterior motives; a person

¹⁵ There are complications if one follows O.E.D. in distinguishing several senses of "generous," such as magnanimous and liberality in giving. I think it more plausible to recognize just one sense: a relative lack of concern for personal possessions, status, advancement, etc., combined with a relatively strong interest in persons or causes not involving one's self directly. If this is correct, then liberal gifts are one manifestation one might expect, given appropriate circumstances.

is not generous if he is disposed frequently to make gifts to establish a favorable public image, avoid income tax to a government he hates, and so on. But what is it to make a gift "without ulterior motives?" Apparently it is to make a gift with such an understanding of the circumstances as not to rule out—or rather so as to imply an important role for—motivation by an *intrinsic* (as it was called above) desire to assist, that is, a desire at least not wholly derivative from a desire to advance one's self in some way. It appears, then, that when we try to describe behavior which is to be expected if a person is generous, we cannot do so without bringing in reference to the person's desires/aversions/motives/goals in some way or other. And *any* kind of behavior which is testimony to the requisite kind of motivation is reason for calling a person generous. This result is precisely the opposite of what one would expect if the Direct Disposition Theory were true, at least for the trait generosity.

There is a further serious difficulty with the Direct Disposition theory, which again should incline us toward the motivation theory. This is the fact of how we make, or revise, trait-estimates in view of single pieces of behavior. Consider an example. Suppose a person, whom we have deemed kindly and sympathetic, does a mean thing. We do *not* then say: "This is one of the cases permitted by the trait of sympathy, since to be sympathetic is only to be disposed to do kindly things on relatively frequent occasions"—as the Direct Disposition theory implies we should. On the contrary, the mean action constitutes a problem for us, and we revise our assessment of the person's character unless we can find some plausible explanation which permits us to reconcile the behavior with a strong aversion to hurting other people. (Sometimes we are satisfied that there *must be* some explanation, although we have no idea what it is.) We do not concede that a sympathetic person, just occasionally and with no strong counter-motivation (or other explanatory circumstance) does a mean thing; we do not think of traits of character as statistical facts. Consider again how we make our original assessment. It is true that information about frequency of behavior is relevant, if we happen to have it; if we know that a person relatively frequently performs services for others, apparently with no ulterior motive, we infer a trait of generosity—and this is consistent with the motivation theory, since relatively frequent behavior of this sort is

testimony to a relatively high motivation of a permanent sort, in normal circumstances. But we also infer traits on the basis of behavior in a single situation, when this can be explained only by the presence of a strong motive presumably of a relatively permanent kind. For instance, if in circumstances of very great temptation to do something else, e.g., acquire the desired property of another when there is no possibility of being detected, one does not do so, one is assigned the trait which consists in desire/aversion which would explain one's behavior. In other words, given our knowledge of the situation and of what may be presumed to be various motives operating, we reconstruct the "psychological field" and motivation of the individual so as to explain his behavior, and where called for we assign a character-trait. A relative frequency is not what we determine before we assign a trait; evidence for assignment of (relatively permanent) wants/aversions is what we do determine.

One other line of possible objection to the motivation theory may be considered in concluding the present section. It may be objected that the motivation theory has some plausibility for traits like generosity, considerateness, etc., but very little for others which are important and universally recognized as traits of character.

It is clear, however, that various other important character-trait-names, not hitherto considered, can be construed naturally along the lines of the motivation theory. Callousness is a lack of sympathy. Honesty is an aversion to deceit and/or the appropriation of the property of other persons. Conscientiousness is an aversion to failure to do one's duty. Unselfishness is a relatively high interest in the welfare of other persons. Kindness is aversion to causing any kind of distress in others. Truthfulness is aversion to deviation from the truth, or to the kind of interpersonal relationship which results when one party indulges in deception.

But possibly other important traits of character cannot easily be so construed. How about courage, temperance, prudence, reliability, and modesty? Let us consider just two of these, courage and temperance, which seem likely to be as difficult cases for the motivation thesis as we shall find.

What then is courage? It is not just a matter of standing (or being disposed to stand) in the face of dangers—for a man might do this because he does not recognize dangers for what they are, or because, in view of his skill or power, they are not really dangers for him. Courage is at least a disposition

to stand up against what the person *thinks* are serious threats of some kind. Courage is not a matter, as Aristotle seems to have thought it partly is, of not feeling fear; for a person who trembles from fear need not be short of courage if he stands his ground, however useless he may be in combat. What courage requires is primarily a negative feature of the system of desires/aversions, *not* setting the highest store by personal safety and position in life. Courage is not, of course, a matter of setting *no* store by these things; that would be foolhardiness in the highest degree. There is something in Aristotle's suggestion that courage requires a desire to maintain one's honor, or an aversion to disgrace; the point is that there are *some* values which have priority over personal safety in a courageous man, whereas a cowardly man will back down on any issue when a serious personal threat arises. Suppose a person has a scale of values, of priorities—involving the obligations of his position, the welfare of other persons, his long-range goals in life—in which presumably personal safety does not outweigh everything else. A person is cowardly, then, in the highest degree if any risk to his personal safety and position will impell him to a course of action divergent from that implied by his scale of values. The more courageous a man is, the less he will be swayed, when something he deems important is at stake, by even serious and imminent threats of personal harm or status. Courage, then, is essentially the absence (the general question of what level or degree is discussed below) of an all-absorbing attachment to personal safety and position.

Let us turn now to temperance or self-control, in the sense which contrasts with self-indulgence. Philosophers have sometimes thought of this as having to do, at least primarily, with the bodily appetites, especially for food, drink, and sex. This is too narrow. A man is self-indulgent if he omits his daily exercises in favor of the morning paper; a student is self-indulgent if he spends time watching football games on television when he should be devoting it to his studies. We can put it generally: a person exhibits temperance or self-control when he foregoes immediately enjoyable experiences which he knows conflict with his long-term welfare, or when he engages in immediately unpleasant activities (like doing push-ups) when his long-term welfare calls for it. Now, what kind of person is apt to behave in this way? The answer seems to be: a person who has a strong aversion to impairing his long-range prospects for a good life.

The self-controlled man is one whose aversion to risking these is sufficiently great to overcome very considerable attractions of immediate enjoyment and the irksomeness of unpleasant activities. His aversion is strong enough so as to bring to mind the relevance of present activities to them, and to control behavior when this relevance has been brought to mind.

IV. TRAITS ARE DISPOSITIONS OPERATIVE IN A NORMAL FRAME OF MIND

So far our (first approximation) proposal has been that what it is for *X* to have a trait of character *T* is for it to be a relatively permanent feature of *X* that, were *X* in situation *S* he would . . . , and were *X* in situation *S'* he would . . . , and . . . ; and furthermore, for these dispositions not to be derivative from other needs/aversions. We have sketched what might replace the '*S*'s, and how the blanks might be filled in, for the case of sympathy.

It is now necessary to complicate this proposal somewhat further. For we do not necessarily withdraw our judgment, for instance, that *X* is sympathetic, just because he fails to respond in the specified way even when his situation satisfies the antecedent of one of these subjunctive conditionals.

The most obvious type of situation in which failure of behavior specified in the consequent does not lead to withdrawal of trait-ascriptions is that of severe emotional disturbance. We do not call a person unsympathetic if he fails to respond sympathetically five minutes after being discharged from his post. We tend to regard this phenomenon as one of the "primitivizing" or "regressive" effects of strong emotion. In this respect character-traits are very like intellectual capacity; a person with a high I.Q. may, in a state of emotion, do very poorly on a standard test. And emotional disturbance may affect the influence any need/aversion may have on action.

The effect of recognition of someone's emotional condition on our ascription of traits to him is, however, rather more complex than the preceding paragraph suggests. There are degrees of emotional upset. A given state of upset will permit some trait-attributions to stand unrevised despite absence of the behavior specified in the relevant consequent, but not necessarily every attribution; in a given state of shock ungenerous behavior might not lead us to withdraw the attribution of

generosity, whereas forging a check in the same state probably would lead us to withdraw the attribution of honesty. Moreover, if we assign a man a trait in a high degree we expect it to manifest itself even in abnormal situations; we expect a highly considerate man not to act in a mean way, virtually irrespective of his emotional situation. And so on.

States of emotion are not the only situations which make us hesitant to revise trait-ascriptions in view of a person's response failing to satisfy the consequent of one of the subjunctive conditionals. Information that the person is intoxicated, or under the influence of tranquilizers, or just extremely fatigued, has the same effect.

We cannot incorporate all such complications into our illustrative account, say, of sympathy. Nor, of course, can we incorporate it into our pattern of analysis, except in some crude way which simply points to the fact that certain complications would have to be taken into account in any complete analysis of a particular trait. One way of doing something to recognize the complication officially I propose to adopt: that of introducing the phrase "if in a normal frame of mind" into the antecedent part of any subjunctive conditional functioning as the partial analysis of a trait-name. Thus we shall insist that the analysis of "is sympathetic" will be of the form, "would feel disturbed, *if he were in a normal frame of mind* and perceived some sentient being to be in acute distress."

V. TRAITS ARE WANTS/AVERSIONS WITH A STANDARD LEVEL OF INTENSITY

For all that has been said so far, a person might be assigned a trait of character if the intensity of the want/aversion which constitutes the trait were as small as you please, anything above indifference. This is surely counterintuitive. We do not call a person "sympathetic" if in the most severe cases of distress he manages only the slightest inclination to relieve it, not enough to motivate substantial expenditure of effort. Obviously some standard level of want/aversion is required for attribution of a trait; we must consider this.

We do not need conclusions about the degree of want/aversion proper for ascription of a trait in its absolute form, for the purpose of comparative statements. It is clear that a person has a given trait in *higher* degree than another, if the data indicate a stronger want/aversion. *X* is *more*

sympathetic than *Y* if he gets *more* disturbed at the sight of distress in another, feels more relief when the distress is removed, would expend more energy in relieving the distress, would feel guiltier about doing nothing to help someone in distress, and so on (all the appropriate exceptions and reservations being taken into account).

Our present concern, however, is with trait-attributions in their *absolute* form: judgments just that a person is honest, conscientious, etc. It should be noticed, however, that we do not really need to be clear about this matter, as distinct from judgments in comparative form, for some of the purposes of moral philosophy mentioned in the introductory remarks; for the purpose of explaining what it is to have a *defect* in some trait of character it is enough if we understand what it is to say that a person has *more* or *less* of a trait than he ought to have. Nevertheless, an understanding of the use of trait-names in their absolute form is of interest.

It is obvious that trait-attributions in their absolute form do not give very precise information. What they do is say that a person *at least barely qualifies* as having a trait, no more. If we want to be a bit more definite, we can say that a person is, say, highly sympathetic, or moderately sympathetic, or perhaps sympathetic enough typically to do a certain kind of thing. What, however, is the minimum to which a person must come up in order to qualify as having a certain trait at all?

One possible proposal is to say that a person qualifies as having a trait, if he has it at least to the degree to which the average person does (or, if we gave tests, came out in the top 60 per cent). But this is implausible, since it forbids us to say that there is any trait of character which everyone or nearly everyone has.

What seems the correct answer may best be explained by an example. Take sympathy. Each of us has a conception of what a sympathetic person will do in a situation of a certain kind. For instance, we may think a sympathetic person will interrupt a friendly tennis game to tend to a child who has fallen off his bicycle. If he won't do that, we do not call him sympathetic. Each of us could give an answer as to whether a person would or would not be reckoned sympathetic by us, if he did or didn't take pains of a certain degree when needed to relieve distress in a given type of situation. This range of situations, and the range of degrees of trouble a person is willing to take, define the least a person has to do to qualify, for a given speaker,

as being sympathetic. What ascription of the term "sympathetic" does is assert that the person's desire/aversion is at a level to produce the required degree of effort in these circumstances, or stronger. We may put this by saying that what a person does, who assigns a certain trait, is affirm that the person has a disposition identical with a relevant desire/aversion up *at least to the standard level*. I suggest this is true of all trait ascriptions.

We must complicate this suggestion slightly. In the first place, we make allowances for how the agent in question perceives the situation. For instance, we do not regard a young lady as unsympathetic if she fails to rush to assist a victim of an accident, if we know that she knows she always faints at the sight of blood. Furthermore, the effort required for one person to do a certain thing may differ greatly from that required for another. A nervous person in a demanding business post may find it much harder to give up cigarettes than does someone else; and the temptation to watch a baseball game on television may be much greater for a young boy than for his little sister. We make such adjustments in deciding whether a trait-name may be applied, but what is expected of a normal person in response to a correct perception of a situation remains a kind of baseline.

An interesting question is whether the criteria which define the lower limit of applicability of a trait-name are roughly the same for speakers of a given language community, or whether they vary

from person to person, or from group to group. There is presumably a limit to personal variation, since the terms are learned mostly by hearing them applied to persons in concrete situations. The conditions of learning do not guarantee uniformity throughout a speech community, however, since terms are learned by hearing their use by a small group, especially the family group. And use by different groups may differ somewhat, particularly in view of the fact that character-trait-names have a flavor of moral praise or blame about them, so that attribution of the term carries a certain moral commitment. Thus business men might not call a man dishonest for failing to report profits on expense accounts for tax purposes, whereas government employees might call him so. So we may expect some variations in the criteria for the lower limit of applicability of some traits.

Thus the criteria for assigning trait-names are probably both somewhat variable and fuzzy at the border-line. We might add that the actual degree of the want/aversion in a given person is probably not perfectly precise, either. There may be some things a person would always do, some he would never do; there may be other things he would do on one day, but not on another. Character traits are possibly somewhat like physical abilities; and athletes have their off days and we cannot predict precisely how well a given high-jumper will do on a given day.

The University of Michigan, Ann Arbor

Received April 7, 1969

III. SUBSTANCE, REALITY, AND THE GREAT, DEAD PHILOSOPHERS

MICHAEL R. AYERS

I. INTRODUCTION

It seems probable that more philosophers are now taking the history of philosophy seriously than has been the case for some time. But if there is one thing radically wrong with our approach to the past, it is the often unquestioned assumption that the first task of the commentator is to isolate from a philosopher's work what is consonant with currently respectable theories, as if the only way of bringing the dead to life is to patch them up, by a kind of cosmetic surgery, as fit participants in some modern debate. It has been overtly maintained by the influential writer with whose interpretations I shall mostly be concerned that the commentator's "dominant" problem is thus to sort the wheat from the chaff. His premiss is equally explicit: "Most of the *Critique of Pure Reason* is *prima facie* dead, because *prima facie* dependent upon wholly indefensible theories."¹

Such an unhistorical attitude, in its extreme manifestation reminiscent of the worst kind of Biblical exegesis, may yet have a little to be said for it: for we may hope that the study of past thought will directly influence and improve our own thinking, so that we avoid the role of those writers on the history of ideas who seem incapable ever of getting to grips with ideas themselves. Nevertheless it is in general worth taking steps to avoid the besetting sin of working philosophers, which is to forget that the history of philosophy, like the history of architecture, is there to be studied in a quite objective way. No serious architect is going to design a building just like a Greek temple, but that is no reason for dismissing such architecture as "dead," or for trying to bring it to life by the desperate pretence that the Greeks shared the aims and ideals of the Bauhaus. It is parochial, not to say philistine, to refuse even to try to understand the alien in its own terms.

But there is another criticism that should weigh even with a philosopher lacking objective curiosity about the history of his own discipline: it is simply false that a training in analytical thought alone qualifies us to penetrate to the kernel of philosophical truth or interest within any historically conditioned husk. The less exhilarating fact is that it is often impossible to get more than a rough and distorted notion of the real meaning of a philosophical work, however deathless it may be, without a reasonable concern for historical context and scholarly comparisons.

In illustration and support of this last assertion, I shall discuss a paper by Jonathan Bennett,² in which, on the basis of purely philosophical arguments and the alluring method of "rational reconstruction," quite false conclusions are drawn about the structure of Berkeley's philosophy and its relation to Locke's. I shall endeavor to show that a more historical approach might have avoided these radical errors.

Bennett's aim is to convict Berkeley of some "bad mistakes" not commonly attributed to him. These are an inability to distinguish the problem of substance from the problem of perception or the external world, the mistake of getting the distinction between primary and secondary qualities mixed up with both these problems, especially with the problem of perception, and a general misreading of Locke on all these issues. Berkeley has escaped condemnation for his errors, Bennett contends, only because commentators have uncritically inherited both his version of Locke and his philosophical confusions.

II. SUBSTANCE, MATTER, AND BERKELEY'S "CONFLATION"

Bennett starts by distinguishing two familiar lines of thought, one leading to what he calls the "Substance Doctrine," the other to the "Veil-of-

¹ Jonathan Bennett, "Strawson on Kant," *The Philosophical Review*, vol. 77 (1968), p. 340.

² "Substance, Reality, and Primary Qualities," *American Philosophical Quarterly*, vol. 2 (1965), pp. 1-17.

Perception Doctrine," both of which find some expression in Locke's *Essay*. The first kind of argument may be illustrated as follows. A quality must, if it exists, be a quality of something. It cannot just exist by itself. So to think of a quality or set of qualities as having existence is to think of it as possessed by, or as inhering in, or as supported by a kind of thing or substance that *can* exist independently. Consequently it is impossible to analyze the concept or "idea" of any kind of thing that can be conceived of as existing by itself, like the idea of gold or of man, merely as a collection of ideas of qualities: the idea of the substance supporting them must be superadded.

A standard comment on this line of thought is that it is unilluminating as an explanation of the existence or coexistence of qualities to postulate something that is unknowable except as their "support," i.e., as what enables them to exist. Locke is himself so sarcastic about this manoeuvre that it might be thought that only a natural bent for inconsistent compromise prevents him from discarding the notion of substance altogether. But it is worth saying now that this would be a serious mistake.

Locke stigmatizes the idea of substance as relative and obscure. It is so because it is merely the idea of whatever supports qualities, relying for all its content on the notion of "supporting;" and it is furthermore obscure because this notion is metaphorical. But Locke never draws, and indeed categorically denies the conclusion that it is meaningless to assert the *existence* of substance. For Ockham's razor plays little or no part in his program of demonstrating that there is a great deal in the world that we do not understand. Locke's scorn is directed against the pretensions of the "Doctrine of Substance" to convey significant information about the *nature* of substance. His main purpose in the *Essay* is to clip the wings of *dogmatic* philosophy, including dogmatic rationalism, and his chief weapon is the principle that all our ideas come from experience. But he never suggests that rationalism is, intrinsically, a flightless bird. He is, in fact, a deliberately doubting or anti-dogmatic rationalist, and the commonly alleged inconsistency between his "Empiricism" and his rationalism is a myth. His attitude to the concept of substance ties in with his rationalism. He holds that the idea of substance performs two significant functions for thought as it ordinarily occurs in combination with a group of simple ideas of qualities and powers: first, it implies that the

designated complex property is instantiated and, secondly, it implies that this complex property has in nature a real, nonconventional unity. When we are led by experience to attribute a number of qualities to the same sort of thing or substance, we imply that such qualities share in this unity, and are linked to the essential properties of the substance by a real bond. Locke tries to use the first of these alleged functions of the idea of substance, its existential import, to account for the difference between *a priori* sciences and the *a posteriori* investigation of nature; and, indeed, for the very possibility, as far as we are concerned, of a comprehensive, nontrivial *a priori* science like mathematics or, as he supposes, ethics. On the other hand, his entirely rationalist conception of the bond that links the qualities of one substance—they flow from its essence as the properties of a triangle flow from its definition—stands behind his belief in the *theoretical* possibility of a demonstrative natural science based on clear and distinct ideas, adequate definitions or "real essences." Consequently, so far from being readily dispensable, the "idea of substance" plays a central role in Locke's philosophy. I shall have more to say about it.

The second train of thought identified by Bennett leads to the so-called Causal or Representative Theory of Perception, the doctrine that the immediate objects of awareness are ideas or sensations, and that these provide us with premisses for a reasonable inference to the existence of an objective world that causes them, which they depict or represent. Locke's attitude is perhaps less ambivalent toward the real world "behind" our sensations than it is, on the surface, toward substance, but he is also less interested in the epistemological problem of perception, e.g., in the objections raised by the sceptic of the senses, than he is in the nature of science and substances.

Bennett's thesis about Berkeley is that he has confused two doctrines or problems, "as different as chalk from cheese," by identifying "substance," the support of qualities or properties, with "reality," the cause and object of sensations. Bennett tries to demonstrate this thesis by quoting passages in which Berkeley is said to be shooting quite indiscriminately at substance and at the external world as if they were one and the same target, welding them together under the title "material substance," "which Berkeley uses lavishly and which hardly occurs in Locke."

The line of Bennett's argument can be briefly

illustrated by his comments on *Principles* I § 17 and § 37. In § 17, after attacking as meaningless the explanation of the concept of "material substance" as "the idea of being in general, together with the relative notion of supporting accidents," Berkeley concludes with the words:

But why should we trouble ourselves any further in discussing this material substratum or support of figure and motion and other sensible qualities? Does it not suppose they have an existence without the mind? And is not this a direct repugnancy and altogether inconceivable?

Then in § 18, he makes use of a quasi-sceptical argument that we could not in any case know of the existence of "solid, figured, moveable substances without the mind." Bennett's comment on all this is that Berkeley "launches off from 'existence without the mind,' etc., into an attack on the veil-of-perception doctrine! In this passage, a complaint against a wrong analysis of subject concepts is jumbled together with a complaint against Locke's insufficiently idealist analysis of reality."

Bennett's comment on § 37 is to the effect that in its context it is wholly ambiguous as between an assertion that a substance is no more than a set of qualities, and an assertion that an external object is nothing over and above the sensations we have of it; and that this ambiguity accurately reflects Berkeley's confusion. The conflation, he says elsewhere, might be embodied in the ambiguous sentence "Things are no more than collections of ideas." Thus Berkeley is incidentally accused of using both "sensible quality" and "idea" ambiguously, each for the other as well as in its usual sense.

III. THE ARGUMENT OF THE PRINCIPLES: THE PREMISS—§1-§6

My counter-thesis is that Berkeley's interlocking remarks about substance and reality are always subordinated to a carefully-structured train of thought. Bennett's neglect of this argument is well illustrated by his "diagnosis" of the origins of Berkeley's "blunder." The conflation of *substance* with *matter* has occurred, he says, because of Berkeley's "double use" of the word "idea" to mean both *sensory state* and *quality*. Bennett's explanation of how this unrecognized ambiguity came about is as follows: the expression "idea of white" would be applied by Berkeley primarily to a sensation or image of white, but because he

follows Locke in holding that to think of the meaning of "white" is to call to mind an image of white, and because it can be said that one who thinks of the meaning of "white" is thinking of the quality whiteness, the term "idea" comes unnoticed to mean *quality* as well as *sensation*. It is the failure to distinguish ideas (=sensations) from ideas (=qualities) that has led to the corresponding failure to distinguish the doctrine that a real world exists behind our sensations from the doctrine that substances exist over and above qualities.

The objection to this diagnosis is that it glides over the attribution to Berkeley as an "underlying assumption," in Bennett's words, the very principle that he is most at pains to demonstrate in the early paragraphs of *Principles* I, the paradoxicality of which he so fully appreciates: the doctrine that the distinction commonly drawn between an idea, or collection of ideas, and its object, whether quality or thing, is unfounded. In § 3-§ 6 Berkeley adduces three arguments against differentiating between ideas and sensible qualities: in § 3 he argues that the *esse* of all sensible objects is *percipi*, and that it is, in virtue of the appropriate meaning of "exists," clearly unintelligible that they should exist out of a mind; in § 4 he argues that sensible objects are indeed perceived, and only ideas can be perceived; in § 5 he accuses the doctrine of abstract ideas of encouraging us to drive an illicit wedge between the sensible object and the sensation or perception of it. Later, in § 8, he argues that, as conceived of specifically by the resemblance theory, the distinction is absurd, because an idea can only be like, or represent, another idea.

Now it might be objected that Berkeley here talks of sensible *objects* or *things* rather than qualities. But in fact he makes it clear that by "sensible object" (and even "thing") he means *primarily* the sensible qualities, e.g., color, taste, heat, smell, and figure. In these paragraphs he deliberately plays down the distinction between *qualities* and *things*, e.g., apples, books, tables, and houses, which he explicitly treats as collections of qualities. His examples and arguments show that he is so well aware of the sensation/quality distinction that he can consciously make it his target, and his clearly expressed reasons for rejecting it have little enough to do with the assumption about meaning to which Bennett attributes the alleged conflation.

It is ironical that Bennett should find "strong confirmation" of his diagnosis in § 78, which he

quotes as "Qualities . . . are nothing else but *sensations* or *ideas*, which exist only in a mind perceiving them." So far from being evidence of an "underlying assumption," this sentence is the restatement of what Berkeley must see as a hard-won conclusion, although the omission of the words "as hath been shown" may tend to disguise the fact.

IV. THE MAIN ARGUMENT—§ 7

Greater difficulties for Bennett's interpretation are raised by § 7:

From what has been said, it follows, there is not any other substance than *spirit*, or that which perceives. But for the fuller proof of this point, let it be considered, the sensible qualities are colour, figure, motion, smell, taste, and such like, that is, the ideas perceived by sense. Now for an idea to exist in an unperceiving thing, is a manifest contradiction; for to have an idea is all one as to perceive: that therefore wherein colour, figure, and the like qualities exist, must perceive them; hence it is clear there can be no unthinking substance or *substratum* of ideas.

First, although Berkeley has been trying to refute the common opinion that "sensible objects have an existence natural or real, distinct from their being perceived," now that he turns his guns on "unthinking substance or substratum" he takes himself to be raising a *fresh* topic. His conclusion is supposed to *follow* from what has been said, but in such a way as to allow "fuller proof." Now this in itself ill fits the accusation that Berkeley talks indiscriminately of objects without the mind, substance, and matter. In fact, his order of exposition has been chosen with considerable care: for, as later appears, while the belief in the independent existence of "sensible objects," the target of the opening paragraphs, is taken by Berkeley to be a popular misconception, the notion of an unthinking substratum he rightly regards as a more sophisticated concept, so that belief in material *substance* is an intellectual sin characteristic of philosophers.

In even greater conflict with Bennett's account is Berkeley's precise statement of the conclusion that he intends to prove more fully: i.e., "there is not any other substance than spirit, or that which perceives." These are hardly the words of someone about to embark on an attack on "substance," which he identifies with "matter"! In fact, Berkeley's target is not substance but dualism, the doctrine that there are two kinds of substance, one of which, body, does not think, i.e., perceive or will.

The reason why Berkeley does not maintain in § 7 that he has already refuted dualism, but does maintain that this refutation *follows* from the preceding paragraphs is because he believes that the conclusion already established—i.e., that the distinction between idea (=sensation) and sensible object (=quality or collection of qualities) is invalid—will serve as the crucial *premiss* in a more complex argument against "unthinking substance." Let me run through this "fuller proof." Color figure, motion, etc., the sensible qualities, are (as has been proved) "ideas" in the mind. To say that an unperceiving thing "has an idea" is a contradiction. Therefore it is a contradiction to say that an unperceiving thing has a color, shape, etc., i.e., that color, shape, etc., exist or inhere in an unperceiving thing. Therefore, the philosophical theory that these qualities inhere in a special kind of nonmental substance is false. On the contrary, they inhere and must inhere in what perceives them. Therefore, spirit is the only kind of substance, for a substance is that in which qualities inhere, and all qualities inhere in spirit.

My presentation of this remarkable argument is a slight expansion, in the light of many other passages, of the characteristically subtle and succinct § 7. Because it is vital to an understanding of Berkeley's thought, I offer the following fuller and more graphic presentation.

Berkeley takes philosophical dualism, i.e., the view that mind and matter are, equally, ultimate realities or substances, to entail the existence of at least four kinds of thing, with certain dependencies between them, indicated here by arrows:

A. Dualism

INTERNAL	EXTERNAL
Spirit ← (Sensations or Ideas)	(Sensible Qualities) → (Unthinking Substance)

He takes it that his previous arguments have ruled out the conception of a "sensible thing or object," such as light and color, heat and cold, extension and figure, distinct from the perception of it. He therefore feels free to treat dualism as follows:

B. Immaterialism

EXISTENT	IMPOSSIBLE & REDUNDANT
Spirit ← (Ideas = Sensible Qualities)	Unthinking Substance

Sensible qualities do indeed have and logically require a dependent status. As our intuition tells us, they must "exist in," "inhere in," "be supported by," and "be had by" something. But what they really depend on is spiritual substance. For, contrary to common opinion, they *are* ideas in the mind. The postulation of a different kind of substance, therefore, to "have" and "support" them, is unnecessary. Such a move is, indeed, self-contradictory, being the postulation of an unthinking owner of ideas, an unperceiving percipient.

It is worth considering two possible reactions to all this. First, it might be objected that external sensible qualities cannot be identified with internal objects of perception. Secondly, it might be objected that whatever reason lies behind the intuition that sensible qualities cannot exist on their own, unsupported or unowned, it is not the same as the reason why an "idea" must be internal to some mind. E.g., it might be said that sensible qualities "depend on" material things because, quite generally, *all* properties are universals, shadows of descriptions, so that they "exist" only by favor of particulars so describable: whereas sensations or "ideas" are dependent on percipients because they are sensory states.

These two objections are at least in order. Their differences from Bennett's misplaced criticisms are informative. Bennett accuses Berkeley of blindly conflating or identifying the *property/thing* relation with the *sensation/external object* relation. But Berkeley is in fact consciously trying to explain away a special case of the *property/thing* relation, i.e., the *sensible quality/material object* relation, by replacing it with the *sensation/percipient* relation, which he claims will explain and justify the subordinate status that we intuitively, if hazily, allot to sensible qualities when we suppose that they need an owner. Moreover, as we shall see, Berkeley is fully aware that the *sensible quality/material object* relation is a *special* case of the *property/thing* (or *property/substance*) relation, as conceived of by his opponents.

Neither of the objections that I have imagined, even if they are correct, can accurately be supposed to convict Berkeley of confusion or conflation. They neither draw distinctions that he failed to notice, nor expose "underlying assumptions" that he failed to question. They merely deny what he supports by argument throughout the *Principles*. For § 7 is the cornerstone of the *Principles*.

³ I am grateful to Rom Harré for some helpful suggestions toward improving an earlier version of this section.

V. PRIMARY AND SECONDARY QUALITIES—§ 9–§ 15³

In § 9 Berkeley introduces both the distinction between primary and secondary qualities and the terms "matter" and "corporeal substance." I quote the whole paragraph, a part of which is the object of Bennett's most powerful scorn:

Some there are who make a distinction betwixt *primary* and *secondary* qualities: by the former, they mean extension, figure, motion, rest, solidity or impenetrability and number: by the latter they denote all other sensible qualities, as colours, sounds, tastes, and so forth. The ideas we have of these they acknowledge not to be the resemblances of any thing existing without the mind or unperceived; but they will have our ideas of the primary qualities to be patterns or images of things which exist without the mind, in an unthinking substance which they call *matter*. By matter therefore we are to understand an inert senseless substance, in which extension, figure, and motion, do actually subsist. But it is evident from what we have already shewn, that extension, figure and motion are only ideas existing in the mind, and that an idea can be like nothing but another idea, and that consequently neither they nor their archetypes can exist in an unperceiving substance. Hence it is plain, that the very notion of what is called *matter* or *corporeal substance*, involves a contradiction in it.

I take Bennett's interpretation of Locke's discussion of primary and secondary qualities, and of Berkeley's view of Locke, to be roughly as follows. Behind Locke's arguments lies the conceptual truth that to call something square is to make a much more extensive and fundamental kind of claim about the world than it is to call it red. The status of, e.g., shape as a "defining" property of body might be illustrated by the fact that, while we can readily imagine a cosmic situation in which there was not sufficient agreement among percipients or coherence among their sensory states to make objective color judgments possible or appropriate, yet if we try to imagine a similar situation with respect to judgments of shape or size we find that our whole concept of the physical world begins to crumble. It is an inessential aspect of Locke's discussion of this truth, so Bennett thinks, that his background assumptions about perception and reality sometimes lead him to express it by saying that ideas of primary qualities are *like* qualities actually in bodies, while secondary qualities exist, except as "bare powers," only in the mind. Yet Berkeley, according to Bennett,

takes Locke's distinction for "a qualification of the veil-of-perception doctrine": i.e., he misreads the remarks in the *Essay* about the subjectivity or relativity of ideas of secondary qualities as if they were intended as a concession to the arguments of scepticism or idealism, and misconstrues the demarcation between the two kinds of quality as Locke's last ditch in the defense of realism. The ground given for this accusation against Berkeley is that he tries to argue for total idealism by first accepting some of Locke's arguments for the subjectivity of ideas of secondary qualities, and then denying the difference between secondary and primary qualities in this respect. This is illicit, Bennett thinks, because these arguments were originally aimed solely at pointing up an important and interesting conceptual distinction that Berkeley could easily have accommodated within immaterialism, and have virtually nothing to do with the case for or against realism—still less with the doctrine of substance.

A weakness of this criticism is that it places too much weight on Berkeley's association of the kind of argument briefly mentioned in Locke's discussion of the water seeming hot to one hand and cold to the other,⁴ with the results of his own quasi-sceptical reflections on the fact that the state, position, etc., of the perceiver help to determine how *any* aspect of the world is perceived. For this association by itself can give no ground for the conclusion that Berkeley actually read Locke's argument as a conscious concession to the sceptic or idealist, and as a move in some debate about the existence of the external world. It is true that, when Hylas appeals to the distinction in the *First Dialogue*, he is clutching at a straw in the midst of such a debate; but it does not follow that Berkeley thought of Locke—or Boyle, or Newton—in just the same way. There is, in fact, positive support from the text for the historically more likely hypothesis that he saw the denial that the ideas of secondary qualities have external patterns as, at most, an unwitting and indirect concession to idealism. For Hylas claims that his admissions about colors, sounds, etc., are "no more than several philosophers maintain, who nevertheless are the farthest imaginable from denying matter." The significance of linking the distinction with a particularly *strong* form of realism will soon be explained.

On the latter half of § 9 Bennett comments:

⁴ "... whereas it is impossible that the same water, if those ideas were really in it, should at the same time be both hot and cold." *Essay* II. viii § 21.

"How can such a farrago as this be understood—how could anyone spell out in plain terms what it is that is being opposed here—except on the basis of an elaborate exposure of the two conflations?" "The two conflations" are Berkeley's "appalling conflation of the question about the appearance/reality distinction with both the question about substance and that about the primary/secondary distinction."

It is easy enough to answer this rhetorical question. Berkeley is opposing the philosophical and scientific conception of body expounded in Locke's *Essay*, but also in the works of other writers, according to which body is defined by a set of primary qualities and is regarded as an active and independent kind of substance, different from spirits finite and infinite. Another philosopher who held such a view is Descartes. It is not important that the essential properties of Descartes' corporeal substance, the attributes of extension, are fewer than Locke's primary qualities; but it may help to explain Berkeley's concentration on extension and motion. Newton is another intended target, and so very probably is Boyle. In fact § 9 introduces a new or more specific target, but no new argument. In § 7 Berkeley has already attacked the general philosophical doctrine that sensible qualities are "had" by unthinking substances. He has contended that sensible qualities plus their substratum constitute, not an unthinking thing, but a perceiving thing, a spirit and its ideas. In § 9 he is merely spelling out the obvious but crucial point that the preceding arguments also conflict with the specific Cartesian, Lockean, and Corpuscularian conception of body as an independent substance.

A possible criticism of § 9, on a point that may have contributed to Bennett's more sweeping denunciations, is that Berkeley confuses the argument by using the term "matter," here and elsewhere, for the postulated unthinking *substrate* of primary qualities rather than, as presumably he should, for the unified defined whole, the extended, moving, solid, etc., substance. But this unclarity is not primarily the fault of Berkeley. There is obviously a genuine difficulty for anyone wanting to express dualism in the form of a doctrine that there are two *kinds* of substance, while at the same time thinking of a substance as something underlying all its properties. A Cartesian would have a reply in the doctrine of the essence

through which a substance can and must be conceived: he would deny that he really thinks of a substance as something other than all its attributes. Locke ascribes such difficulty to our ignorance of essences—to the fact that “we have no idea of what [substance] is, but only a confused, obscure one of what it does.”⁶ For Berkeley, the difficulty arises as a difficulty in expressing what he opposes. Nevertheless it is clear enough that he is opposing the doctrine that there is a kind of substance that has the property of extension and lacks the property of thought.

Why should Berkeley take the trouble to identify as a specific target the concept of material substance embodying the primary/secondary distinction? Bennett’s explanation is simply that Berkeley regarded the doctrine of primary and secondary qualities as a weak point, at which he mounts his attack on Locke because he mistakenly thinks that Locke’s realism is here already on the run in the face of idealist arguments. Now it may be true that Berkeley makes some sort of attempt to extract a debating advantage from the “acknowledgment” that secondary qualities exist only in the mind—Berkeley enjoyed irony as well as Locke—but there are several better reasons why he should take notice of the distinction.

One of these, not perhaps very important, is an advantage that is independent of the acceptance of any particular argument of Locke’s. Even if, as I believe to be the case, he rejected *all* Locke’s arguments, it would be worth while pointing out that any appeal to ordinary intuitions about the idea/quality distinction will apply equally against the official, express doctrines of his chief philosophical rivals. This fully explains a number of allusions to the primary/secondary distinction, e.g., in § 46 and § 49, where, as in § 9, it is quite beside his purpose to mention any *argument* for it.

Far more important is the connection between the distinction and rationalistic dualism. Bennett presents his own view of “what is interesting in the distinction,” and concludes that Locke is “struggling to express and defend” this truth. But such a conclusion has no justification. Extension may be, in Bennett’s sense, a “defining” property of physical objects, as color is not. But if such is the conceptual truth behind the distinction, then Locke is trying to express something other than this truth—some-

thing a good deal more like what he actually manages to say.

Let us, however, turn first to the crystalline reasoning of Descartes. Who could deny that the central doctrines in Descartes’ philosophy are associated with his belief that extension is really in body, while color exists only in the mind or is “related simply to the intimate union that exists between body and mind”?⁶

Some of Descartes’ discussion of secondary qualities may have an intrinsic philosophical interest, apart from its relation to the rest of his metaphysics: e.g., the argument according to which hardness, exemplifying inessential qualities, may be shown to be inessential from the conceivability of a world in which hardness is never actually felt.⁷ But it would be absurd to pretend to interpret Descartes’ own intentions in this discussion without relating it to his doctrines about substance or, indeed, to his rationalist epistemology in general. Take, for example, the following contrast:

We have experience of [*scil.*, mechanical change] not just by one sense, but by several, by touch, sight and hearing; we also distinctly imagine and understand it. This cannot be said of other things that come under our senses, such as colours, sounds and the like, which are perceived . . . by single senses; for their images are always confused in our minds, nor do we know what they are.⁸

This cannot be understood except in the light of the earlier, long discussion of clear and distinct ideas and the claim made there that we have a clear and distinct conception of “a corporeal or extended nature that may be moved, divided, etc.,” but not, as yet, of the causes of the sensations of pain, color, taste, etc. It would be particularly absurd to allege that Descartes’ contentions about the unequal status of primary and secondary qualities could be expressed within phenomenalism. He is most emphatically concerned to put extension “out there” as an attribute of body, the self-sufficient subject-matter of physics, (an intelligible substance requiring for its existence only God’s concurrence) separated from the various sensations excited in the mind by the motions of body.

There is not space here for a full justification of the view that Locke’s interest in the distinction

⁶ *Ibid.*, II. xiii. §§ 18f.

⁷ *Principles of Philosophy* II. 3.

⁸ *Ibid.*, II. 4.

⁹ *Ibid.*, IV. 200.

between primary and secondary qualities, and *ipso facto* his interest in substance, is a good deal more Cartesian than is commonly taught. There are too many levels in his rationalistic semi-scepticism or anti-dogmatism for any easy exposition of his theory. Briefly, Locke regards it as a fundamental implication of the attribution of a set of qualities to a sort of substance (i.e., by adding the general idea of substance to the complex idea of those qualities) that they all "flow from the particular internal constitution or unknown essence of that substance;"⁹ so that, if we had adequate ideas of substances, i.e., of their real essences, we should have nontrivial knowledge of necessary connections holding between their properties. This is as clearly stated as anywhere, perhaps, in II. xxxi. § 6, a most important passage for understanding Locke's rationalism. I shall quote a small part of it. Locke is arguing that, contrary to common (and Peripatetic) opinion, our ideas of substances are not adequate, i.e., do not correspond to real essences:

... if you demand what those real essences are, it is plain men are ignorant and know them not. . . . Such a complex idea [as we ever have] cannot be the real essence of any substance; for then the properties we discover in that body would depend on that complex idea and be deducible from it, and their necessary connexion with it be known, as all properties of a triangle depend on and . . . are deducible from the complex idea of three lines including a space.

Many passages commend or presuppose the same model for scientific knowledge. But Locke also tells us that, although we lack adequate ideas of particular physical substances, chiefly because of our ignorance of their minute parts, so that "a perfect science of natural bodies (not to mention spiritual beings)" is beyond us,¹⁰ yet the corpuscular hypothesis *approaches* this rationalistic ideal, affording some kind of general intelligibility, and mitigation of our ignorance of connections within the physical world.¹¹ This advantage derives, apparently, from the alleged necessary connection between felt resistance and the communication of motion by impulse.¹² Apart from these and many other more or less explicit pro-

nouncements, Locke's often expressed, quasi-Cartesian belief in the special intelligibility of mechanical explanations is enough to suggest that the primary/secondary distinction might very well be related to a conception of body as an ontologically independent substance, capable of being the subject of a demonstrative physics even if we are not fully capable of being demonstrative physicists. In all, there is quite enough in the *Essay* to justify Berkeley in seeing the primary/secondary distinction as an integral part of rationalistic dualism. We know from his notebooks that he paid special attention to a crucial passage.¹³ But in any case we know that he associated the distinction directly with the Cartesians; they are mentioned eight times in connection with it in the notebooks, whereas Locke is mentioned only three times.¹⁴

It might be objected to my interpretation of Locke that his treatment of the primary/secondary distinction follows not Descartes but the Corpuscularians who are also mentioned by Berkeley in the notebooks: i.e., the more cautious tradition of Gassendi, Boyle, Newton, and Clarke, who do not associate their concept of the essential properties of material substance with such a very rationalistic concept of substance. Locke's close connection with this perhaps less metaphysical, but still dualistic tradition is undeniable, but I believe that he owes much to Descartes: the *Essay* consciously links the two traditions. He may attack Cartesian and Scholastic dogmatism and defend empirical method, but the basis for both attack and defense is an acceptance of the Cartesian ideal for scientific knowledge and understanding, together with the contention that our present, perhaps irremediable lack of adequate ideas of substances and their essences makes this ideal unattainable. I strongly suspect that the desire to conjoin these two influential philosophies of science is one of the most important motives for Locke's philosophy, if we are to grasp its underlying unity.

Yet even if *Essay* II. viii is *simply* an expression of straight corpuscular doctrine, this is far from being the conceptual insight ascribed to Locke by Bennett. For even on that narrow interpretation, what Locke is "struggling to express" and defend would not be a conceptual truth about different

⁹ *Op. cit.*, II. xxiii. § 3.

¹⁰ *Ibid.*, IV. iii. § 29.

¹¹ *V.*, e.g., *ibid.*, IV. iii. § 16.

¹² *V.*, e.g., *ibid.*, IV. iii. § 14. Cf. II. iv.

¹³ *V. Philosophical Commentaries*, 533, Luce's edition, in A. A. Luce and T. E. Jessop (eds.) *The Works of George Berkeley* (Camdon, N.Y., 1948).

¹⁴ Noted by Luce, *Berkeley and Malebranche* (Oxford, 1967, 2nd imp.), p. 62.

kinds of sensible quality, but a now discarded scientific conception of matter. The relativity argument against the existence of external patterns of our ideas of secondary qualities (i.e., the argument implied in the first sentence of II. viii. § 21, which almost certainly derives from Galileo) might then be seen as a bad argument for the desired conclusion—whatever conceptual truth can be seen lurking “behind” it. The rejection of this one argument is not especially damaging to Locke’s case, since the real point of this paragraph is simply to illustrate the possibility and reasonableness of a scientific explanation, in terms of constant mechanical processes, of the order and coherence of ideas of secondary qualities. Certainly one aim of the whole chapter—whether or not I am right about its function as *prolegomena* to the rationalist theory expounded later on in the *Essay*—is to draw out the implications for his doctrine of “ideas” of the *physics* that Locke accepts. That is why he apologizes for engaging in “physical enquiries.”¹⁵

Berkeley had a very straightforward motive for making a particular target of the primary-qualified matter of contemporary physics: for unlike ordinary men (or ontologically coherent phenomenologists) the Corpuscularians undoubtedly conceived of the primary qualities of the minute parts of matter as the most real or basic properties of the world we see and touch, although they are imperceptible. Even more importantly, they represented matter as a *causally* self-sufficient substance: the explanation and source of all physical activity is to be found in the primary-quality essences. On Berkeley’s view, an ordinary (pious) man might be expected readily to allow that God really causes all sensible change, and that talk of physical causation is a *façon de parler*. But the physicists thought that real causal explanation, real scientific understanding, is achieved specifically by penetration into the essences of physical substances. Hence Berkeley would see the primary/secondary distinction as going hard in hand with a sophisticated doctrine that matter is a substance *qua* agent or source of activity: in § 9 he is not, as Bennett seems to think, evincing a misreading of a single, rather dreary argument hardly more than implicit in Locke, but is intending to identify the form of materialism most dangerous to religion, the target of § 101–§ 117. The word “inert” stands as a subtle harbinger of this later, climactic denunciation of Newton and the physicists who deify matter by

making it a cause and, therefore, an independent substance.

An understanding of all these reasons for Berkeley’s estimation of the doctrine of primary and secondary qualities makes it easier to question the assumption that he tries to extract from Locke’s arguments about secondary qualities not only a lever against Locke himself but also a proof that all qualities exist only in the mind. For Berkeley’s published writings contain no serious attempt to apply such a lever or present such a proof. There is not space for a full justification of this contention, but the careful reader of *Principles I* will notice that not until § 14 does Berkeley associate his immediately previous argument from the relativity of determinations of some primary qualities with the arguments of “modern philosophers” about secondary qualities. The latter are said to apply “with equal force” to primary qualities. Yet this careful phrase is followed by an explicit denial that they disprove the externality of *any* quality:

“... this method of arguing doth not so much prove that there is no extension or colour in an outward object, as that we do not know by sense which is the true extension or colour of the object.”¹⁶

The argument that Bennett attributes to Berkeley is not to be found in the *Principles*.

Bennett’s hypothesis that Berkeley misread Locke on primary and secondary qualities is made to explain not only the course of the argument when Berkeley is directly concerned with the distinction, but also the many glancing references to extension, motion, and solidity: e.g., even the sentence in § 16, “It is said extension is a mode or accident of matter, and that matter is the substratum that supports it,” excites the rebuke that “‘extension’ . . . has nothing in particular to do with substratum-substance but does have to do with primary qualities.” In fact, such references to particular qualities generally serve merely to designate the *kind* of substance under criticism. Sometimes, however, there is a special point in the quality mentioned. This is true of § 16: the references to extension in the opening sentences quoted by Bennett are taken up in the last sentence, “It is evident *support* cannot here be taken in its usual or literal sense. . . .” Berkeley is delicately referring to the kind of difficulty described in the *First Dialogue* as “the peculiar difficulty there must be, in conceiving a material substance, prior to and distinct from

¹⁵ Cf. Bennett, *op. cit.*, pp. 15ff.

¹⁶ Cf. *Philosophical Commentaries*, 265.

extension, to be the *substratum* of extension." The difficulty arises, of course, because in their literal senses "substance," "substratum," and "support" already involve extension.¹⁷

VI. THE ATTACK ON THE UNINTELLIGIBLE SUPPORT—§ 16—§ 17, § 27, etc.

§ 16 But let us examine a little the received opinion. It is said extension is a mode or accident of matter, and that matter is the substratum that supports it. Now I desire that you would explain what is meant by matter's *supporting* extension . . . if you have any meaning at all, you must at least have a relative idea of matter . . . you must be supposed to know what relation it bears to accidents It is evident *support* cannot here be taken in its usual or literal sense . . . in what sense therefore must it be taken?

§ 17 If we inquire into what the most accurate philosophers declare themselves to mean by *material substance*; we shall find them acknowledge, they have no other meaning annexed to those sounds, but the idea of being in general, together with the relative notion of its supporting accidents. The general idea of being appeareth to me the most abstract and incomprehensible of all other. . . . So that when I consider the two parts or branches which make the signification of the words *material substance*, I am convinced there is no distinct meaning annexed to them. But why should we trouble ourselves any farther, in discussing this *material substratum* or support of figure and motion, and other sensible qualities? Does it not suppose they have an existence without the mind? And is not this a direct repugnancy, and altogether inconceivable?

These paragraphs can only be understood in the light of § 7, § 9, and much else. Their function is to present one side of the contrast between mind and matter, and between the intelligible relation of spiritual substance to its ideas, and the unintelligible relation of material substance to the external qualities supposed to belong to it. Berkeley is not denying that sensible qualities, e.g. extension, require and have a support. He is presenting part of his case that there is no meaningful way of conceiving this requirement except as the necessity that an idea should exist in some perceiving mind.

Previously Berkeley has argued that the concept of corporeal substance is self-contradictory. Now he argues that it is empty and inexplicable. To suppose an *unperceiving* support makes both the support and the supporting totally mysterious, by prescinding the concept of "supporting" from anything known in experience. The last three

sentences of § 17, to which Bennett objects, remind us of Berkeley's view that sensible qualities, being ideas, must be perceived by a spirit anyway. They afford both an allusion to the other difficulties raised by the doctrine of unthinking substance and a resolution, in immaterialism, of the present difficulty over the meaning of "support." Incidentally, the relative intelligibility of spiritual and corporeal substance is a topic in Locke, who sometimes tries to boost his dualism by the consideration that we have as little conception of one as of the other. This is precisely what Berkeley is denying.

Spirit, according to Berkeley, is an intelligible substance because its essence is open to experience. Its relation to sensible qualities or ideas, comprising perception and will are intelligible, although not the object of ideas. This part of Berkeley's theory ought to be well known from § 26—§ 28, where he explicitly draws a contrast between his own account of this understanding as consisting in notions "grounded in experience," and a rival account of mind that brings in an *unintelligible* notion of supporting, in connection with perceiving and willing. After stating that we do not perceive spirit or have an idea of it, he continues:

If any man shall doubt of the truth of what is here delivered, let him but reflect . . . whether he hath ideas of two principle powers, marked by the names *will* and *understanding*, distinct from each other as well as from a third idea of substance or being in general, with a relative notion of its supporting or being the subject of the aforesaid powers, which is signified by the name *soul* or *spirit*. This is what some hold . . . (§ 27.)

This again, out of context, could look like an objection to all talk of "substance" or "supporting." Yet how could it be? Berkeley has just categorically stated that spirit is an incorporeal, active substance, applying to it a traditional definition, "one simple, undivided, active being." His view, roughly speaking, is surely this: The orthodox account of both material and immaterial substance involves the confused and unexplained notion of a support of accidents. In the case of mental substance this notion can be made clear and intelligible by identifying "supporting" with perceiving. In the case of corporeal substance, no such clarification is possible. For the notion of corporeal substance is intrinsically unclear and founded on a muddle.

¹⁷ Cf. F. Copleston, *History of Philosophy*, vol. V (New York, 1959), p. 223.

A clear allusion to the general argument that I am attributing to Berkeley is contained in § 135 and § 136, where he explains the popular (and Lockean) belief that spirit is unintelligible by the fact that we lack an idea of it. Referring back to § 27, he adds:

A spirit has been shown to be the only substance or support, wherein the unthinking beings or ideas can exist: but that this *substance* which supports or perceives ideas should itself be an *idea* or like an *idea* is evidently absurd.

This passage, unintelligible on Bennett's interpretation of Berkeley's argument, is an attempt positively to exploit the impossibility of an "idea" of substance, in order to reconcile us to the lack of an idea of spirit. It contains an overt identification of "supporting" with the perceiving by which he has tried to explain it. In § 136 he virtually declares that "soul" and "substance" are synonymous: "I believe no body will say, that what he means by the terms *soul* and *substance*, is only some particular sort of idea or sensation."

Throughout the *Principles*, the contrast between the intelligible, perceiving, immaterial support of sensible qualities or ideas, and the unnecessarily postulated, unintelligible, impossible, unthinking material support, is elegantly balanced by a similar contrast between the intelligible causality of the will and the unthinking causality postulated by materialists. Just as mind alone can fulfill the substantial function of a "support," so also is it the only source of activity. Berkeley implies, indeed, that "an agent subsisting by itself" is the *definition* of soul and substance alike.¹⁸ In this conception of substance he is deliberately following the letter, but hardly the spirit, of a strong, indeed the only philosophical tradition, stemming from Aristotle, according to which ontological and causal priority go together. Of course, he rejects the usual explanation of activity by essences, in favor, ultimately, of the explanation peculiarly appropriate to voluntary acts, by reference to the conscious purposes of the agent.

I shall not now discuss the *Three Dialogues*, or the many other passages in the *Principles* that help to clarify Berkeley's argument and support my view of it. I must, however, mention *Principles I* § 73, which is a compressed rehearsal of the main argu-

ment, with Berkeley's three targets identified in turn. First the conception of independent, external sensible qualities; second, the *ad hoc* philosophical postulate of the unthinking *substratum*; third, the more technical concept of body, defined as by Descartes and the Corpuscularians. Bennett pounces on the opening clauses of § 74, "Matter was thought of only for the sake of supporting accidents," as proof of the alleged conflation of *substance* with *matter*. But in context Berkeley's thought is quite clearly this: materialists, e.g., Cartesians (the passage occurs in a discussion of Malebranche), must concede that they talk of extended substance in *addition* to thinking substance because they believe that extension does not inhere in thinking substance and that it must inhere in, or be supported by, some substance. The grounds for dualism are removed, when it is realized that, while the second belief is true, the first belief is false. Extension, and every other sensible quality, is and must be supported by a substance, i.e., by thinking substance, as idea in perceiver.

VII. CONCLUSIONS AND MORAL

Let me now summarize my interpretation of Berkeley. His central and coordinating principle, and the conclusion that he wishes to retain after every concession is made to the objectivity and reality of the sensible world, is that mind is the only kind of substance, both *qua* independent support of whatever needs a support in order to exist, and *qua* active principle or cause of whatever needs a cause. Sensible qualities, as our intuitions tell us, need both. Ideas need both.¹⁹ The central argument consists in an elaborate and careful contrast between mind and matter, with reference to their eligibility for the status of *support* and for the status of *cause*, to their intelligibility, and to the reasons that can be adduced for asserting their existence or anything about their nature—their knowability. A first move in this argument, but only a first move, is the "proof" that sensible objects are not external to the mind, and that sensible qualities are not distinct from ideas; from which Berkeley can argue, for example, that the need that a sensible quality has for a support is to be identified with, and explained as, the need

¹⁸ §§ 135–139.

¹⁹ Hume evidently understood Berkeley, as we must if we are to understand Hume (*cf. Treatise I*, iv. 5.: "the curious reasoners concerning the material or immaterial substances, in which they suppose our perceptions to inhere," "Inhesion in something is supposed to be requisite to support the existence of our perceptions," etc.).

that an idea has for a perceiver. For Berkeley, as for Descartes, the treatment of the "problem of perception" stands very much subordinate to a more complex main theme. Like the Aristotelians, and like Descartes, Malebranche, Spinoza, Leibniz, and Locke, Berkeley makes his metaphysics and philosophy of science hinge on a conception of substance. Like some of these, he attacks rival doctrines about substance. Like Locke in particular he is liable to be misread as if he objected to the concept of substance *per se*; although his language makes it plain that he is not in fact doing so. An important, for Berkeley no doubt the important part of his theory is the status of God as a super-substance that, in the words of the *Dialogues*, "produces and supports" the sensible world.

If I am right, then it is quite inadequate to see Berkeley simply as a philosopher concerned with the "problem of perception,"²⁰ who does not stay the course of phenomenalism because he is too interested in the prospect that certain problems for the phenomenalist about causation and objectivity require or justify recourse to theism. Nor is Berkeley's account of matter as satisfactorily summed up in the dictum "*esse is percipi*," as it is often supposed. Indeed the role that the principle that *esse is percipi* or *percipere* plays in Berkeley's philosophy cannot be understood unless it is recognized as a deliberate attempt to parallel Aristotle's doctrine of categories of being, dependent (or secondary) and independent (or primary) existence. This can be supported from his notebooks; as it can be proved that when Berkeley there writes with such excitement of his new "Principle," he is not talking about the *esse is percipi*, as Luce and others assume, but about the principle that mind is the substance that supports sensible qualities, by perceiving them. Hardly any of Berkeley's argu-

ments are fully intelligible without reference to this main argument. Of all the discussions that I have read of his notorious claim that a sensible object cannot be conceived of "otherwise than in a mind perceiving it," I cannot recall one that has recognized what would have been obvious to his contemporaries, that Berkeley is here applying the accepted test for a substance, which Spinoza expresses in his definition of a substance as "that the conception of which does not need the conception of another thing." Hence this passage is not just one quaint argument for phenomenalism that appealed to Berkeley in a weak moment, but an essential move in his attempt to persuade philosophically sophisticated readers that sensible things are not substances.

How then does it happen that the commonly accepted interpretation of a short, lucid, truly magnificent work like the *Principles*, enormously popular as a subject for study, remains highly inaccurate, incoherent, and unbalanced. I suggest that it is not the "casualness" of which Bennett complains, so much as the direction of their own philosophical interest that underlies the treatment that most, if not all commentators have meted out to Berkeley. Consciously or unconsciously they look for what they already conceive of as philosophically important. A tradition of raking in his ashes for phenomenalist ore has quite obscured the outlines of a precise, astonishingly comprehensive and coherent metaphysical argument, only a part of which I have examined here. Recognition of a simple historical fact might have helped to prevent such distortion: in talking of substance Berkeley was employing perhaps the most important, sophisticated, and carefully defined concept in contemporary metaphysics and philosophy of science. It is incredible that he did so unawares.

Wadham College, Oxford

Received November 11, 1968

²⁰ Cf. G. J. Warnock, *Berkeley* (Baltimore, Md., 1953), p. 204.

IV. PLATO AND ARISTOTLE ON BELIEF, HABIT, AND *AKRASIA*

AMELIE RORTY

SINCE we have had several centuries of practice producing examples to show the absurdity of the claim that "To know the good is to do the good," it might be interesting to see whether any interpretations of it are defensible. Tracing the considerations that led as careful a philosopher as Plato to advance such an implausible claim may also clarify ambiguities in the Aristotelian arguments against conventional interpretations of the thesis.

The problem of weakness of character (*akrasia*) is the problem of explaining how it is possible for men to act, intentionally, in a way which is nevertheless contrary to what they know or judge to be the better course of action. Although Plato's formulation suggests that *akrasia* is primarily a problem about the connection between knowledge and action, it is not essential to construe the problem in this way. Even when the issue is formulated in terms of the agent's beliefs or judgments, rather than his knowledge, *akrasia* remains problematic. Before tracing some of the tactics used by philosophers to explain (or explain away) *akrasia*, it should be said that the phenomena present a problem only if one accepts a particular kind of motivational psychology, according to which agents act, other things being equal, or unless they are prevented, so as to maximize what they judge or believe, on the whole, to be their benefit. This psychological premiss may, but need not, take the form of hedonistic egoism. It can also, more plausibly, take the form of a theory of motivation that makes enlightened self-interest the primary motive or "spring" of action. Psychological egoism can, of course, be quite a broad theory, depending on how the limits of the ego are defined.

Solutions to the problem of *akrasia* fall, very roughly, into two categories, those which focus

primarily on the analysis of the effects of belief on action and those which concentrate primarily on the connection between desire and action. Traditionally, the Platonic solution has been interpreted as falling within the first category, the Aristotelian within the second. My primary intention in this paper is to examine these two general types of solutions, to determine what considerations lead a philosopher to adopt one rather than the other, and to assess their respective advantages and disadvantages.

I

The most publicized version of the Platonic solution to the problem of *akrasia* involves a number of distinctions between potential knowledge and its active exercise. This part of the Platonic picture focuses sharply on the relation between belief and action, and leaves the relation of desire and appetition to belief and action blurred in the background. According to the conventional accounts, Plato's solution to the problem of *akrasia* is that since knowledge and belief about what is valuable are always action-guiding, any putative knowledge or belief that fails to guide action must also fail to qualify as genuine knowledge or belief.¹ The man who suffers from *akrasia* only appears to know, or mistakenly thinks that he knows, but obviously cannot actually fully realize or understand what is to his interest. This aspect of Plato's treatment of the problem of *akrasia* is usually interpreted as a denial of the existence of genuine cases of *akrasia*. But the characters in the dialogues describe genuine cases of *akrasia*. Cephalus approvingly cites Sophocles' description of the advantages of age: it brings relief from sexual passion that gives the young no mercy, even when they judge that

¹ This line of interpretation is presented, by W. Jaeger, *Paideia*, vol. 2, II, 64-5; J. Gould, *Development of Plato's Ethics*, p. 6; T. Gomprez, *Greek Thinkers*, vol. II, p. 67. Although many scholars have distinguished the Socratic from the Platonic analysis of *akrasia*, I have coalesced them here for the sake of making the contrast with Aristotle clearer.

acting from such desires is not always advantageous. In describing cases in which *epithumiai* can drag a man like a slave or a wild beast, Plato does not say that if a man cannot follow his *logos*, he has not reasoned at all.² It is more accurate to say that Plato redescribes the phenomena of *akrasia*. By not accusing Plato of simple blindness, we are invited to analyze the rationale for the redescription.

Plato constructed his motivational psychology with a moral intention. He wanted (and I think he was not satisfied that he had achieved his purpose) to give an account of the "ideal" relation between belief and desires, an account that would explain and justify the possibility of educating a man's desires by an appeal to rational self-interest. Such a theory would not only make the conative elements of the soul cognitive, but also make the cognitive functions conative—capable of directing and perhaps even motivating action without additional intervention. Plato, at one time, at any rate, hoped to show that if a man develops his rational capacities, he will simultaneously discover his true desires; his actions will then become free of conflict. He also hoped to show the converse: allowing the full and natural development of a man's desires simultaneously leads to the development of rationality and a sense of proportion. This visionary program of moral education is possible only in a society which has not already permanently corrupted the desires or crippled the rationality of its citizens. It is a moral vision, furthermore, that is well-founded and actualizable only if a certain psychology, a complex theory of the connection between reasons and motives is valid: *The Republic* is, among other things, a sketch of the social and political conditions in which this moral program can be effective; many of the later dialogues represent Plato's attempt to develop the epistemological and metaphysical conditions that must obtain if the Socratic enterprise is not to be illusory, that is, if a rational

examination of the good for men can actually make men better, and if men bent on pursuing their self-interest will eventually see that it coincides with rational virtue.³

Bambrough has pointed out that this enterprise not only presupposes the validity of a particular motivational psychology, but it also involves changing the ordinary use of Greek expressions.⁴ Plato finds himself forced to redescribe moral phenomena. These redescriptions are justified partly by Plato's belief that the revised descriptions are somehow more correct than the old usage; they bring ordinary speech and ordinary belief closer to describing possibilities that are both ideal and in principle actualizable. Plato thinks that it is possible to improve social institutions and individual practices, bringing them closer to what is ideally possible, by bringing about a reform in ordinary speech and opinion.

These are, I think, the sorts of maneuvers that take place in the discussions with Thrasymachus and Glaucon in *The Republic*, and also what is involved in the analysis of *eros* in *The Symposium* and *The Phaedrus*. Plato does not make his tactics explicit, except by trying to "ground" them in an appropriate epistemology and metaphysics, because doing so would undermine his didactic and moral purposes. If he were to make these tactics explicit, he might be open to the charge of sophistry, a charge that would in one sense be justified, but in another be totally misleading.

Not even Socrates can carry on a discussion with anyone whose assumptions are radically different from his own. Generally, he does in fact share at least some of the premisses accepted by his interlocutors. In *The Republic*, he seems to grant, for example, Thrasymachus' connection between virtue and power, and tries to show that vulgar justice is more advantageous than injustice. The course of his argument is intended to modify Thrasymachus' conception of what is truly ad-

² *The Republic* 329 c-e; 440b; 441b; *The Phaedrus* 237d-238d.

³ In *The Theaetetus*, for example, especially the section on Protagoras (161B-168C; 169D-179C), Plato analyzes the conditions for the falsity of perceptual judgments (*aesthesis*) and more generally judgments about matters of fact (*doxa*), opinions expressed in the form "It seems to me that . . ." The moral enterprise requires that although no one desires what is harmful, men may desire an object because they are mistaken about its qualities or mistaken about the consequences of possessing it. Even the abstract arguments about being and non-being in *The Sophist*, necessary to an understanding of the similarities and differences between the sophist and the philosopher, lay the epistemological and metaphysical groundwork for the Socratic program. The relevance of *The Statesman*, concerned as it is with the question of whether a philosopher can transform a bad state, and of *The Philebus*, which is concerned with the question of whether there is an objective measurement of pleasures, are more obvious.

⁴ Renford Bambrough, "Socratic Paradox," *The Philosophical Quarterly*, vol. 10 (1960), pp. 289-300. Here Bambrough shows the relevance of *The Cratylus* to the problem of justifying the Socratic program. In that dialogue, questions about how far it is possible to transform the meanings of words, and questions about the way in which ordinary speech carries philosophical as well as ordinary opinions are discussed (obscurely and without resolution).

vantageous as well as his conception of justice.⁵ In *The Symposium*, he appears to accept the premiss that *eros* is a primary and non-rational moving force, but his interpretations of this premiss differ from those of his interlocutors. He is frequently not so much concerned to establish the falsity of his interlocutors' views as to urge on them a new interpretation of these views. This is of course one of the reasons it is so difficult to interpret the dialogues, not to mention teach them: Socrates is simultaneously evaluating and disambiguating the definitions and hypotheses advanced by his interlocutors. The two processes involved in the *elenchus*, and in any dialectical inquiry—that of disambiguating hypotheses and testing them—are not clearly separable, partly for didactic reasons and partly for philosophic reasons having to do with Plato's belief that hypotheses cannot be tested individually, but only systemically.

When Socrates is trying to persuade his interlocutors to consider their assumptions in a new light, he often engages them in what appears to be a "means-ends" discussion, granting their claims about what they say they find desirable as ends, and then trying to establish that a means different from the one they propose would satisfy these ends more efficiently and fully. This is the form of the argument with Thrasymachus. The claim that virtue is (or is defined by) power is not attacked directly; instead knowledge (*episteme*) is argued to be necessary to secure and to exercise power. If knowledge is necessary to power so that it is impossible to have power without knowledge, then the definition of power must make that dependence clear and explicit. In this way, the definition of the end has been subtly refined if not actually modified. The definition of knowledge is of course also affected by this argument; Plato is laying the ground for his claim that knowledge is a form of power, which, under ideal circumstances, can directly motivate action. Sometimes Socrates employs a variant of this argument, trying to show the interlocutor that an object different from the one he supposes himself to

desire would, in the long run, be more satisfying. The alternative object turns out to modify, refine, or even change the nature of the desire to be satisfied. For obvious didactic reasons, this consequence is generally left implicit.

This technique in argumentation is integrally connected to the doctrine that a desire is partially defined by an agent's beliefs about the object of his desire, and may be radically transformed by a change in the agent's description of what it is about the object that he desired, even when the object of the new desire remains extensionally the same. Thus, for example, someone who desires to seduce Phaedrus may gradually come to see that what he really desires is not to possess Phaedrus sexually but rather to possess and retain the good and the beautiful characteristics of Phaedrus, or the good and the beautiful exemplified by Phaedrus. The revised description of the object of desire changes the nature of the desire, so that the actions that would best satisfy it differ from those the agent originally thought appropriate. Sometimes the argument proceeds in the opposite direction: Socrates tries to persuade the interlocutor that the nature of a sexual desire is a desire for the total possession of an object so that it can be permanently satisfying. He then tries to show that this desire can in fact only be satisfied by redescribing the object desired.

Now what has all this to do with the problem of *akrasia*? To see the connection, let us turn to Santas' careful analysis of the Platonic solution of the problem. According to Santas, Plato requires a distinction between an agent's desires and his evaluations. An agent can act, intentionally, against his better judgment because, while an agent must always act according to what emerges, on balance, as his strongest desire, this desire does not always accord with the agent's evaluation of the action and its consequences.⁶ This solution, Santas argues, has the merit of giving us a criterion for measuring an agent's evaluative rankings independently of his desires and actions. One discovers an agent's rational evaluations by dialecti-

⁵ Many commentators, (e.g. David Sachs in "A Fallacy in Plato's *Republic*," *The Philosophical Review*, vol. 72 [1963], pp. 141-158) have claimed that Plato has committed a fallacy in this argument. This would be true if one took him to be simply giving a straightforward analysis. But, as I shall argue, he is simultaneously engaged in conceptual and moral reform. Many of the paradoxes result from the playful contrast between the old vulgar usage and the proposed Platonic revision. Pointing to a program does not automatically locate a fallacy. One would still have to show why the program is ill-conceived. It is this more fundamental sort of criticism that Gregory Vlastos makes so well in his paper, "Metaphysical Paradox," *Proceedings and Addresses of the American Philosophical Association*, vol. 39 (1965-66), esp. pp. 18-19.

⁶ Gerasimos Santas, "Plato's *Protagoras* and Explanations of Weakness," *The Philosophical Review*, vol. 75 (1966), esp. pp. 27, 29-30. See also his "The Socratic Paradoxes," *The Philosophical Review*, vol. 73 (1964), pp. 147-164.

cal discussion, his desires by observing his actions. But while Santas' distinction does present a temporary solution to the problem of *akrasia*, it could not be Plato's solution. For the Platonic position requires not only that agents act following their strongest desires but also that they desire what they judge to be good for them. In any case, Santas' solution relocates the problem in a new place, as a problem about how an agent's desires can conflict with his judgments about what is best for him.

Plato uses two words that are sometimes differently translated as "appetite" or "desire." To determine whether desires can conflict with rational judgments, we must examine the relation between *epithumia*—the faculty of appetite—and *boulesis*, a cognitive desire.⁷ This terminology brings us to the dynamics of soul and the relations among its various functions. What are we to make of the relation between *epithumia* and specific *boulesei*s? Plato seems undecided whether the distinction between *epithumia* and *boulesis* is (1) a distinction between a non-cognitive instinctual need and a cognitive desire for objects falling under specific descriptions or (2) a distinction between a rash and vague instinctual or impulsive but nevertheless cognitive desire on the one hand and a desire that has been clearly specified in detail on the other. The first alternative makes *epithumia* and *boulesis* different in kind. It is closer to the tri-partite division of the psyche and fits the purposes for which it was constructed. It preserves the independence of appetites and reason. The second alternative makes *epithumia* and *boulesis* different in degree rather than kind; it fits Plato's program as a moral reformer and his beliefs about the rational educability of desires.

The notion of a totally undefined and undirected

appetition is, for a human being, a limiting case; one can barely imagine what it would be. To the extent that we are at all aware of our instinctual desires, we have already specified them and described them as desires of—or desires for. We have beliefs about the sorts of objects and activities that best satisfy them. Such desires are cognitively defined, though still perhaps in a very general way, as for example, desire for sexual satisfaction rather than a desire to have an affair with Phaedrus. Before such desires can direct an agent to any *particular* relevant action, they are capable of, and indeed require, further specification, both in being directed to some objects rather than others, and in being interpreted in one way rather than another, as requiring some sorts of actions rather than others (e.g., philosophical inquiry rather than seduction). As we have seen, these two further modes of specification are not entirely independent of one another since, for many desires, specifying an object affects the nature of the desire and vice versa.

Santas has suggested that the difference between an *epithumia* and a *boulesis* is that the object of every *epithumia* is pleasure whereas the object of every *boulesis* is a good, or at least what the agent takes to be a good.⁸ When Plato speaks as if *epithumia* and *boulesis* are different in kind rather than degree, he seems to suggest something close to what Santas says; but this is a corollary of the primary difference between them. In a sense, Santas is not mistaken in saying that the object of every *epithumia* is a pleasure (though it would be more accurate to say that it is the satisfaction of a want or need) rather than the objects that give pleasure or satisfaction. *Epithumiai* can be thought of as instinctual appetites that become defined and specified as conceptualized desires;

⁷ There is considerable disagreement among scholars about whether Plato intends to draw a sharp distinction between *epithumia* and *boulesis*; even those commentators who argue that Plato does intend to differentiate them disagree about the significance of the distinction. (See, for example, Santas' criticism of the way Croiset and Bodin render the terms in *The Meno*, and his citation of Bluck's interpretation.) I shall be speaking of the distinction as more firmly formed than I believe it was. There are many contexts, it seems to me, where Plato uses the terms almost interchangeably. To argue that he intends, in such contexts, to make fine distinctions seems to me to maul the text to no purpose. Nevertheless, there are contexts where a subtle difference between *epithumia* and *boulesis* is suggested though not fully developed. It is this implicit suggestion of the difference between a physically based need or want, and a conceptually formulated desire, that I am interested in developing. I cannot point to a text where Plato makes that distinction explicitly, but it is implicit in his usage, even in those contexts where he glosses over the differences between the terms. After all, in English, the words "need" and "want" have a loose as well as a narrow sense. We may speak of wanting or needing things that, under cross questioning, we perfectly well realize we only desire or wish to have. It is significant that when Plato is discussing physical needs, such as sex or thirst, *epithumia* is the word he always uses. (Cf. *The Symposium* 204d 5–7; *The Republic* 437b–c; *The Philebus* 20d–22c.) In *The Meno*, during the discussion of whether men can desire what is bad, knowing it to be bad, the terminology shifts from *epithumia* to *boulesis*, at just the point where there is a realization that such desires are in principle corrigible precisely to the extent that they are cognitively defined (*The Meno* 77b–78b). This is the distinction between thought-independent and thought-dependent desires drawn by Stuart Hampshire in *The Freedom of the Individual* (New York, 1965), pp. 47–48.

⁸ Santas, *loc. cit.*, esp. ft. 15.

in their preconceptual form, they have no future tense indicators. An active instinctual need requires immediate satisfaction; it cannot by itself weigh its claims against the long range consequences of its satisfaction. (Cf. *The Republic*, pp. 437-439.) Furthermore, such *epithumiai* are general rather than specific: when one is thirsty what one wants is a drink, whether it is spring water or Melian wine doesn't matter.

The rational evaluation of desires not only places them in a hierarchy; by placing desires in a temporal perspective, it can also at its best redefine a desire in such a way as to make it less insistent for present satisfaction. When *thumos* and *epithumia* lead *logos* astray, it is primarily because the agent is overcome by outrage at a wrong done to him, or by the immediacy of pleasures or the insistence of a need. But a *boulesis* in judgmental form, a desire for an object under a specific description in the belief that it will be satisfying, can place an intense need for satisfaction in its proper temporal perspective, without necessarily denying the desirability of what is desired.

Returning now to Santas' account of Plato's solution to the problem of moral weakness, we can see how there might indeed be a conflict between an *epithumia* and an evaluation, if *epithumia* and *boulesis* are different in kind. But it is not at all clear how there could be a conflict between a *boulesis* and an evaluation, for Plato holds that one *must* desire what one judges to be good. Yet cases of *akrasia* are said not to reduce to simple moral conflicts or conflicts of desires. What complicates the matter further is that many desires are specifications of instinctual needs, directing them to some sorts of objects rather than others. On this interpretation, a *boulesis* could conflict only with its corresponding original *epithumia* by being specified in a way which fails to satisfy the original need. An apparent conflict between an *epithumia* and a *boulesis* would involve a clash between a general instinctive *present* desire for an object that may be lost if not satisfied *now* and a cognitive, rationally evaluated desire for something more specific. A *boulesis* frequently involves a plan of action, and often requires postponement. This is not of course to suggest that a *boulesis* is always true or correct; quite on the contrary, *bouleseis*, are often false or mistaken. Being cognitive in form, they make specific predictions for future satisfactions based on past experience, and are intellectually corrigible in ways in which

epithumiai are not. *Epithumiai* are simply present; they can be educated and improved, but not by purely intellectual processes. The examples Plato gives of *epithumiai*, thirst and sexual desire, suggest that they are the sort of physiological needs that recur periodically—they are wants in the etymological sense: lacks that require replenishment. There is some suggestion that *eros*, at least, provides the energy, or the raw material, for directed and specified *bouleseis*. Only when knowledge or true belief informs desires is it likely that their satisfaction will in fact genuinely satisfy the original *epithumia* of which they are a specification; that is, only when the belief that forms the cognitive part of a *boulesis* is true, is it likely that acting on it will produce genuine satisfaction.

But these remarks give us only general formulae, directions in which we may find solutions, rather than the detailed working out of these solutions. There are still a number of problems, perhaps even dilemmas, that Plato must face. It is still not clear how desires that are partly formed by an agent's conception of the objects that best satisfy them can be different specifications of the same original *epithumia* even though their objects are quite different. The desire for immortality conceived as the desire for procreation is different from the desire for immortality conceived as a desire for knowledge of eternal objects; the two desires would lead to very different sorts of actions. How can they then, if they are quite different, both be specifications of the same instinctual appetite, one of the fundamental *epithumiai* of the human psyche? Evaluating the relative importance of these two desires would require measuring the success of their respective actions in gratifying the original *epithumia* of which they were specifications. It is not at all clear how this could be done without presupposing the original identity of the *epithumiai*.

If the difference between *boulesis* and *epithumia* is a difference of degree and if actions motivated by thought-dependent desires can be given temporal indicators, the *akrates* is ignorant of his interest, or doesn't really desire what he thought he did. This account of *akrasia* explains away counter-examples by redescribing them as cases of deception of some sort. It also represents a retreat from the view that appetite is independent of reason. The problem of *akrasia* reduces to the problem of how someone with innate knowledge of the forms and an innate longing for what is good, could nevertheless make a mistake. It takes the form of asking how someone who really potentially knows

what is to his benefit could nevertheless act, being mistaken about what he really desires. Plato sometimes seems to believe that if anyone who holds an erroneous judgment would pursue his inquiry far enough, he would see that he doesn't really believe what he thought he believed. Similarly he would also argue that if anyone pursues an inquiry into what he desires long enough, he would discover that he doesn't really desire some of the things he thought he had desired. Ideally he should cease desiring them.

But if *epithumia* and *bouleis* are radically different, so that a conflict between an *epithumia* and *bouleis* is not simply a conflict about the priorities of rational desires, then cases of *akrasia* can be construed, following Santas' suggestion, as the victory of *epithumia* over *logos*. Relocating the problem of *akrasia* as falling in the area where there is a conflict between an evaluation (*logos*) and a desire rather than in the area where there is a conflict between an action and a rational evaluation, still leaves us with serious difficulties in explaining the educability of desires, whether they be *epithumiai* or *bouleiseis*. Plato was faced by a dilemma: the kind of psychology he adopted to account for the rational educability of desires, and the continuous cognitive specification of such desires, leads him to the solution often interpreted as the denial of the phenomena of *akrasia*. On the other hand, the kind of psychology that most easily accounts for moral conflicts without redescribing them runs into difficulties with the psychological egoism based on enlightened self-interest that can most easily support the program of moral reform.

What has increased Plato's slipperiness in this matter is that he has assimilated the intentionality of *bouleiseis* with their being formed and defined by beliefs. For Plato, to say that desires are intentional is to say that they are caused by beliefs and are intellectually corrigible. Yet in different contexts, when he wants to show that someone can have false beliefs about his own desires, Plato finds himself needing to distinguish a man's being aware of his desires from his having formed his desires on the basis of his beliefs.

II

I want now to turn to Aristotle's discussion of *akrasia*, to trace his criticism of one version of the Platonic position, to see how he can also be con-

strued as supplying flesh for the schematic and puzzling account of the relation between *epithumia*, *bouleis*, and the agent's beliefs about the objects which best satisfy them.⁹

Commentators on Aristotle's views on *akrasia* tend to stress one of two possible aspects of Aristotle's doctrine. Some emphasize the function of practical wisdom (*phronesis*) in moral action and so are led to think of moral failures as primarily intellectual, though to be sure, failures involving the practical rather than the theoretical intellect. On this view Aristotle's position approaches one version of the Platonic position. But one may also, as other commentators have, emphasize the function of habit in moral action. These commentators argue that the discussion of *akrasia* is given such prominence in the *Ethics* because it is a crucial example of a flaw of character falling within the realm of the voluntary, and is thus central to any analysis of morality. As the title of his book indicates, Aristotle holds that an analysis of virtue must include an account of the development of *ethos*. It is the connections between a man's character, his habits, his emotions, and his knowledge that Aristotle is concerned to explore. The faculty psychology that haunts Plato's account, and that forces him into the dilemmas we have outlined, is all but dropped by Aristotle. He deliberately blurs some of the distinctions implicit in Plato's terminology and yet at the same time adopts some of that terminology to make distinctions of his own.

Aristotle draws a number of distinctions between various types of knowledge. The first set are adapted from some Platonic views: they cluster around the difference between potential knowledge and its active exercise. The second distinction is an Aristotelian innovation; it is the distinction between practical reason (*dianoia praktike*), which is always action directing, and theoretical reason and *episteme*, which are not. (*N.E.* VI chaps. 5 and 6.) Before giving a very general account of Aristotle's analysis of *akrasia*, I want to sketch Aristotle's use of these distinctions. Aristotle follows Plato in distinguishing knowledge that is actively exercised from the implicit or potential knowledge that an agent may have without using it or being aware of having it. This distinction is not to be equated with that between real and apparent knowledge nor with that between true belief and knowledge. Although

⁹ Two excellent books on this subject have recently appeared: Ronald Milo, *Aristotle on Practical Knowledge and Weakness of Will* (Hague, 1966); and James Walsh, *Aristotle's Conception of Moral Weakness* (New York, 1963).

this distinction resembles Plato's distinction between innate potential knowledge and the actualization of such knowledge, it in no way presupposes innate intellectual knowledge of the good.

Aristotle distinguishes several occasions in which potential knowledge or belief fails to be actualized. (*N.E.* 1147a 10-25.) The first, which need not concern us, are cases in which an agent possesses knowledge he is not actively using because he is not in a situation in which such knowledge is relevant. Secondly, potential knowledge may fail to be exercised because it is not yet genuinely and fully possessed, so that its active exercise in appropriate situations is not yet fully habitual. Aristotle does not discuss this sort of case in detail, presumably because one hesitates to call these clear-cut cases of knowledge or belief. Finally, an agent may have knowledge relevant to a situation in which he actually does find himself and yet, for some reason or other, be unable to summon or exercise his potential knowledge. It is this sort of failure that concerns us in cases of *akrasia*.

The second type of distinction Aristotle introduces in his treatment of *akrasia* is that between practical and theoretical reasoning. Practical reason, unlike scientific knowledge, is always connected with the faculty of desire, and its exercise is always directed toward a specific end involving an activity. It is also said to differ from theoretical reasoning in dealing primarily with particular rather than universal truths. Both theoretical and practical reasoning involve a number of distinct but mutually supporting talents or habits. Theoretical reasoning requires the capacity to perform induction, that of apprehending the universal in particular, that of finding the middle premiss in deductive syllogisms, and so on. Practical reasoning also requires a number of distinguishable but related intellectual talents or habits. A virtuous man must have a variety of intellectual abilities: the capacity for deliberation as well as the capacity to choose good ends; practical wisdom (*phronesis*), the temper of mind which selects ends wisely, must supplement understanding (*synesis*); good sense (*gnome*), and cleverness (*deinotes*). (*N.E.* VI chaps. 5-11.)

It is as a combination of desire and thought, Aristotle says (*N.E.* 1139-b 3-5) that a man is capable of action. The expressions Aristotle uses in this quotation are "*oretikos nous*" and "*orexis dianoetike*," which may be translated respectively as "thought motivated by desire" and "desire

directed by thought" without the implication that this involves an interaction between two independent faculties. "*Orexis*" generally is translated indifferently as appetite and desire, although the original connotation is much closer to "a reaching out for . . . , a penchant for" The implication is that the objects of wants are determinate and are not modified or specified by the agent's beliefs about them. Aristotle frequently uses *orexis* in the loose sense, referring indifferently to appetites and desires, as he also sometimes uses *boulesis* and *epithumia* interchangeably. At other times, when he speaks more strictly, he distinguishes the terms: *orexis* is the generic term for appetite, of which *thumos*, *epithumia*, and *boulesis* are species. *Boulesis* is the type of *orexis* that is formed by beliefs that can be true or false; it represents a wish or a desire that a state of affairs should obtain, or a desire for an object under a specific description. It is an attitude toward propositions. *Epithumia* in its stricter sense is the arational form of *orexis*; it is the arational moving principle of the psyche. (*D.A.* 433a 25-28; 414b 2-15.) *Orexeis* may conflict, and this happens when *logos* and *epithumia* are opposed (*D.A.* 433b 5-12). This can occur only in creatures who have a sense of time, for *epithumia* looks only to the present and is incapable of considering the future. There are two kinds of *epithumiai*: those shared by all men (such as hunger and thirst) and those that are particular to some individuals, such as the appetite for a specific food. (*N.E.* 1118b 10-15.) The latter can shade into *bouleseis*. That they already seem—in humans at any rate—to involve an element of cognition or judgment is suggested by the etymological connection to *boule*, plan, counsel, intention. But one does not take counsel or deliberate about ends: one desires or wishes for them, always taking them to be good. (*N.E.* 113a 15-20).

With these distinctions in mind, we can now turn to Aristotle's account of *akrasia*. There are many different sorts of moral failures, not all of them attributable to ignorance of one's welfare or true desires. It is sound practical wisdom rather than scientific or theoretical knowledge that is a necessary but not a sufficient condition for the development of moral virtue. The question of whether *akrasia* is *primarily* due to ignorance or to some other defect is moot, since Aristotle is listing independent conditions, all of which are necessary and none sufficient for moral virtue. All sorts of habits must supplement that of reasoning well about practical matters. The virtuous man must

not only have good intellectual habits of various sorts, but also sound habits of action involving habitual desires of genuinely good ends that are properly described, as well as good emotional dispositions of various sorts. One does not become moral simply by learning to reason as the moral man reasons; one must also acquire the habit of acting as the moral man acts, from deliberation, on the basis of one's active knowledge, and not merely after deliberation, "repeating one's knowledge as an actor repeats his lines." And of course the deliberation must be directed at achieving ends that are not only genuinely beneficial, but that are also desired in the right way under relevant descriptions.¹⁰

Since this sort of education involves the simultaneous training of action-guiding desires, emotions, the capacity for deliberation, and the acquisition of practical wisdom, and since these subtly reinforce one another, moral education can go wrong in a variety of ways. If one lives in a corrupt society, or has been badly educated, or has an unfortunate temperament, one can come to accept as desirable ends that are in fact bad. This is what has happened to the wicked man. But one can also be brought up by those who dishonor deliberation, or one can have a rash temperament, and so either fail to develop good deliberative habits, or else not acquire the habit of acting in accordance with one's deliberations. The man who suffers from *akrasia* may deliberate well about ends that are desirable, but he cannot be relied on to act primarily from his deliberations about what is truly good. Aristotle thinks that this happens primarily in the sorts of situations that involve the pleasures of taste and touch, eating and sex; but he allows that it can also happen—less culpably—in situations involving an agent's pride. In all other situations, the *akrates* may act according to his practical judgment, following his conceptions of his self-interest.

But our account is by no means complete yet. Aristotle himself does not say, but he might well grant, that there are several sorts of *akrasia*, involving different sorts of failure to follow one's better knowledge. Some cases of *akrasia* are primarily cases of moral inertia, in which a man finds the effort of doing what he judges desirable too

strenuous. In such cases, failure to act may harm no one except the agent, and may harm him only slightly. These types of *akrasia* involve deficiency, and are often misclassified as self-indulgent acts. It would be implausible to say that the agent is, in such situations, overcome by passion or desire; on the contrary, his desire for what he takes to be to his interest is insufficiently strong to overcome habits of lassitude. The self-indulgent man, unlike the *akrates*, is acting according to his principles, rather than against them. He simply has the wrong sorts of principles.

There are also what we may call cases of *akrasia* involving excesses. It is these sorts of cases that Aristotle has in mind when he says that the man suffering from *akrasia* is overcome by passion so that he cannot use, but only mouths, his better judgment. Austin has pointed out that such actions are not necessarily slobbering, rash, or animal-like. They may be done elegantly, slowly, with finesse, and even deliberately. Furthermore, a man who does something against his better judgment need not do that particular action habitually. He need not habitually seduce students to do so intentionally yet against his better judgment; nor need he have an irascible disposition to act intentionally from rage and pride, and yet against his better judgment. An *akrates* is not vacillating; he is torn. He is the type of man who habitually has difficulty in denying or postponing the satisfaction of certain sorts of desires and impulses. Unlike the self-indulgent man, he does not satisfy them on principle but rather against his principles.

This raises the question whether *akrasia* is essentially a flaw of character or whether a man who is not an *akrates* may, on a single occasion, suffer from *akrasia*. Aristotle's answer to this question would be parallel to his discussion of virtues and vices. Virtue and vice are primarily attributes of character; from one virtuous or vicious action one cannot determine a man's state of character (just as one cannot determine whether a man really speaks a foreign language from hearing him utter one sentence in that language). Seeing someone act against what he claims to be his better judgment in a single case does not really tell us why he did it, whether he

¹⁰ I am here glossing over thorny problems about which Aristotelian scholars have debated long and carefully—problems concerning the relations among *phronesis*, virtue, *proairesis*, and voluntary action—because I do not think they essentially affect the analysis of *akrasia*. And although many commentators have concentrated their attention on the practical syllogism, I think we would be better off explaining the cryptic passages on the practical syllogism in the context of Aristotle's general theory of action, rather than struggling to illuminate the obscure by the more obscure.

really did judge another course of action to have been better, or whether given the opportunity he would repeat his action. We may say of someone who on a single occasion acts against what he claims to be his better judgment that he seems to have acted as an *akrates* does, without committing ourselves to a diagnosis of his character. Until we know whether he is likely to behave in a similar fashion on similar occasions, we cannot say whether this particular action counts as a case of *akrasia* for this agent. The question of whether someone who is not an *akrates* could on occasion suffer from *akrasia* is no more and no less problematic than the question whether a virtuous man could perform a vicious act, or a vicious man could perform a virtuous act. Aristotle's answer is—typically—that in a sense he can, and in a sense, he cannot. Actions out of character do of course occur, and they are explicable, though the details of the explanation may involve a long narrative falling far short of scientific rigor or validity.

Cases of *akrasia* may also be classified according to the reasons for the agent's inability to exercise potential knowledge relevant to the situation. Such knowledge may be difficult to actualize because it almost fails to count as knowledge at all. These are the cases Plato sometimes takes as paradigmatic. For Aristotle, they are really only borderline cases of *akrasia*: an agent may have learned fashionable formulae even though he in fact does not even really believe or perhaps even fully understand them. Such cases differ from those in which the inability to actualize relevant genuine potential knowledge is not a failure of knowledge, but a failure of character, a failure in exercising the habit of acting directly from the agent's deliberation about what, all things considered, is best.

This may sometimes happen because an agent is not clear about his preferences; it may also happen when an agent acts upon beliefs he has revised, but which have already formed highly developed and strong habits. When an agent faces a conflict of desires, he cannot solve his problem simply by assigning desires their relative strengths and then (straightway) acting so as to maximize the satisfaction of those desires that have the highest priority. A desire of greater strength but low degree of habituation may be set against one of relatively low strength but high degree habituation. Active knowledge of minor importance may be set against important but relatively inactive knowledge. Failure to muster all one's talents and habits (*hexeis*) in this complex and difficult evaluative

process may sometimes be a failure of the practical intellect, but it is not always one. Though the normal moral agent is not responsible for being clever, he is responsible for developing what practical sagacity he has and using it actively and relevantly. *Akrasia* is considered a flaw of character rather than simply an intellectual flaw because it involves habits of action and emotion as well as habits of mind. In particular, it involves the habit of acting from knowledge, following all the deliberations that have been performed, even in situations where it is difficult to do so, rather than following one set of desires that have great immediate strength.

This account of *akrasia* does not, I think, violate Aristotle's view that actions are motivated by an agent's desires, by his conception of his self-interest. A man suffering from *akrasia* may have developed a weak character before he could realize that it would be damaging to him. When he realizes and regrets his inability to act directly from his knowledge, he is unable to change because his weak habits are so deeply entrenched that they form part of his character. According to this interpretation, *akrasia* involves a special kind of second-order habit, a habit of actualizing other sorts of habits, or a habit of making action-guiding knowledge actually guide action. Much more needs to be said about such second-order habits, to show that postulating them does not simply generate an infinite regress permitting the problem of *akrasia* to reappear on a higher level as a problem about how a man can develop a character that is detrimental to his interests. Though many first-order habits are simply acquired, others are acquired by choice. Are the habits which explain *akrasia* all acquired before a man has been able to realize that they are damaging? Are all second-order habits character traits and does the account of their development differ from the account of the acquisition of first-order habits?

A man's character, his *ethos*, is largely formed by his habits; but, unlike many lower level habits, it is not something that he can change. It is, furthermore, much more generalized; it affects the way other habits are used, for good or for bad ends, skillfully or unskillfully. This, however, is also true of many lower level habits: they are difficult if not actually impossible to change, and they can be exercised in different ways, for good or ill, skillfully or unskillfully. The question arises whether the difference between a man's character and his habits is a difference in order and kind,

or whether it is a difference of degree. While these problems seem exactly parallel to those Plato faces about the continuity and discontinuity of *epithumia* and *boulesis*, they do not, I believe, present Aristotle with a similar dilemma. Since Aristotle is not committed to providing an epistemological foundation for a moral program, he need not take these alternatives as forced options, but can say, as Plato cannot (without jeopardizing his moral program), that in a sense character and habits are continuous, and in a sense they are not. Some types of men find it easier to change their habits than others.

It might be remarked that if this is Aristotle's account of *akrasia*, it is not an explanation but a loose genetic description. To say that *akrasia* is a failure of character and habit is not to explain the phenomenon but only to classify it. This may be right, but it is no argument against Aristotle. The Kantian task of "explaining the possibility of the phenomena" is not one Aristotle set himself in ethics; indeed he thought it a misguided enterprise. The reasons for a man's suffering from *akrasia* will differ in every case. The best that can be done by way of explanation is to point in the direction where the details will be found, to give as it were, their general longitude and latitude.

Rather than speculating about Aristotle's solutions to these problems, let me begin putting order into these observations by suggesting that the difference between Plato's and Aristotle's motivational psychology and their respective accounts of the phenomena of *akrasia* result in part from the differences in their approach to ethics and meta-ethics.

Since Aristotle has no vested interest in explaining the possibility of an intellectual re-education of desires, he is not committed to constructing an ideal model for a psychology that explains how it might be possible for a man to become moral, free of conflicts damaging to himself and to society, by engaging fairly late in his life in a dialectical discussion with Socrates. He does not have to face the question of whether anyone who reasons well about practical matters would in the long run be led to desire the proper ends. Aristotle regards the phenomena of moral education as much too com-

plex for that question to be answered with a simple "yes" or "no." Since he hasn't committed himself to an educational program, he is not required to construct a theory to explain the ideal case.

The problems that interested Aristotle in ethics were primarily of an analytic rather than a pedagogical kind. He wanted to describe the variety of factors that lead us to classify certain types of action as voluntary, to analyze the manifold skills exercised in different virtues, and thus to describe the character of a well balanced man who functions reasonably well.¹¹ According to Aristotle, the rational education of desires that Plato describes is something that can indeed take place, but only under special conditions, when a man's habits (*hexeis*) have been properly formed, when he has a good character, a sound constitution, and a fairly good education. This is not to say that a man can become virtuous only if he is virtuous, but that he is only capable of becoming virtuous if he has not been hopelessly ill bred and ill educated. A sound character and constitution by no means assure virtue, for the acquisition of virtuous habits still requires patient practice as well as a certain amount of good fortune.

The Platonic dialogues exhibit but neither demonstrate nor even defend Aristotle's conclusions. The arguments in the dialogues attempt to make a more ambitious case than is warranted by the dramas they portray. It is very rare—if it happens at all—that an interlocutor is convinced by an intellectual argument in such a way that we can believe he has been truly changed by it. Some interpreters of Plato (J. Klein, for example, in *A Commentary on the Meno*) argue that it is the reader or the bystander who is meant to be the real interlocutor: it is he who is to be improved. One can share Plato's scepticism about the force of the written word (even—or perhaps especially—in dialogue form) in effecting such changes. Still, despite his scepticism about the success of the enterprise, Plato seems to have thought it worthwhile to do what can, within severe limits, be done. Though it would be hard to say that any interlocutors become virtuous in the course of their discussions with Socrates, some characters—Callicles and perhaps even Meno—are shaken in

¹¹ It is of course true that there is a strong Platonic—one might even say neo-Platonic—strand in Aristotle's ethics. He is not only interested in analyzing the elements of virtue, but also in ranking types of temperaments and lives according to the degree to which they fulfill human potentialities. I have not discussed Aristotle's view of the superiority of the contemplative to the practical and productive lives, because these views are not systematically connected to any program of moral reform. Though he describes the psychological and political conditions for the contemplative life, and though he discusses the rhetorical and political techniques for effecting social change, Aristotle does not construct his epistemology to account for the possibility of such change.

their dogmatism and acquire some awareness of their ignorance. If a radical improvement occurs at all, it happens to very young men who, like Theaetetus, may be presumed to have reasonably good habits of mind and desire, or at least not to be hardened in bad ones. Nor can a man, even if he is young and of good character, become virtuous simply by engaging in an intellectual conversation with Socrates; to benefit from conversation with Socrates, it is necessary to love him. Plato hoped to indicate under what conditions the rational education of desires is possible. Initially he hoped to show how these political, social, and psychological conditions could be brought into being. But the dialogues show that this type of change in the conditions themselves is only theoretically, only ideally, possible, and even then, only within a limited range. The practice he portrays and describes only warrants a theory as sceptical as Aristotle's.

III

But what, one may well ask, is the truth of the matter? It is all very well to say that a philosopher's conceptions of morality and motivation will affect his explanation of *akrasia*, and to claim that descriptions of *akrasia* vary because philosophers approach the problem from entirely different perspectives. It may well be true, furthermore, that there are varieties of *akrasia*, all falling in the borderline region where we have some reasons for classifying an action as intentional and even deliberate and yet also have reasons for classifying it as a species of involuntary action. But as Donald Davidson says: "We ourselves show a certain weakness as philosophers if we do not go on to ask: Does every case of incontinence involve one of the shadow zones where we want both to apply and to withhold some mental predicate? Does it never happen that I have an unclouded, unwavering judgment that my action is not for the best of all things considered, and yet where the action I do perform had nothing of compulsion or the compulsive? There is no proving that such actions exist; but it seems to me absolutely certain that they do. And if this is so, no amount of attention to subtle borderline bits of behavior will do anything to resolve our central problem."¹²

I want to sketch an attempt at fulfilling that obligation. How can we plausibly finish the following schema for a story? "In doing this action, X

suffered from *akrasia*. Though he acted intentionally, he did so against his better judgment. He chose what he judged on the basis of one line of reasoning to be *prima facie* more desirable than its alternative, instead of doing what he judged, on the basis of a different line of reasoning to be, *all things considered*, the better of the two alternatives. He would not have done this action if _____." It seems to me that there are two ways in which sad stories of this sort can plausibly be finished. (1) "X would have acted differently if he had thought more carefully about the matter and fully realized the consequences of his actions." When we finish the story in this way, we come down softly on "He *knew* better," suggesting that the agent didn't really know better, that he was unaware of the full details of the consequences of his action. When we end the story in this way, we look to Plato for an account of the failure, and find ourselves distinguishing the consequences of merely entertaining a judgment hypothetically, "mouthing it," from the consequences of genuinely accepting a judgment, perhaps vividly knowing and believing it. It is possible to consider a judgment abstractly, hypothetically accepting it, "believing it for the sake of the argument," and gradually come to suppose, mistakenly, that one has fully accepted and understood it. An agent really believes a judgment when he is not only prepared to entertain its logical consequences for the sake of the argument or for the sake of appearances, but when, other things being equal, he is also ready to act on it whenever the occasion arises unless he is prevented from doing so.

(2) "He would have acted differently if his actions were less irrational, that is, if he were in the habit of following his deliberations about what, all things considered, is the better course, instead of simply doing what seems desirable under one line of reasoning." When we finish the story in this way, we come down softly on "He could have done otherwise." In such cases, an agent who fails to act on his considered judgment about what is (on balance) best, seems not to have developed the habits of following his reason. In these situations, when we are often at a loss to decide whether the action, though still voluntary, comes from a strongly entrenched habit of some sort, we cover our uncertainty by classifying the action as a case of *akrasia*. In some sense it seems clear that the agent could have done otherwise; when we are not sure whether an action may be in some mar-

¹² Donald Davidson, "How Is Weakness of the Will Possible?" in Joel Feinberg (ed.), *Moral Concepts* (Oxford, 1969).

ginal sense involuntary, it does not by any means follow that we think it compulsive. It is perfectly possible for an action to be involuntary in one sense and voluntary in another. The difficulties of determining whether to withhold or apply this predicate are not merely legendary—they beset us, in a practical way, at every turn. When we end the story in this way, we look to Aristotle for an account of what may have happened. Sometimes—depending on the agent and the situation—the first ending of the story seems most plausible; sometimes both endings seem right.¹³

I believe that all actions called instances of *akrasia* do fall into one or both of these areas. If this means that I deny that there are genuine cases of *akrasia*, then I suppose that I do deny them. But if this counts as denying the existence of *akrasia*, then I confess I do not know of anyone, including Aristotle, who does not in this very loose sense deny them. Even Davidson, who is bent on saving the phenomena, classifies *akrasia* as a type of irrationality. The *akrates* has violated what Davidson has called the principle of continence: “perform the action judged best on the basis of all available

reasoning.” And although Davidson warns us against asking why anyone would act on less than all the available evidence he has, it is not clear why this question is forbidden.¹⁴ Presumably asking it simply reiterates the original question on a new level, as a problem about why anyone should suffer from *akrasia*, knowing it to be irrational. But no one who judges that irrational actions are harmful would intentionally act irrationally except from some sort of ignorance or character flaw.

My attempt to fulfill the philosophical obligation of taking the existence of *akrasia* seriously and explaining its possibility seems to have failed, for I am always pushed to those borderline regions Davidson describes. I am tempted to think that this happens because the obligation cannot be fulfilled, and the question we should ask is not “How are cases of *akrasia* possible?” but rather “Under what conditions do people tend to suffer from *akrasia*?” If these questions have a psychological rather than a Kantian flavor, so much the better.¹⁵

Livingston College,
Rutgers University

Received January 28, 1969

¹³ A. P. Mourelatos suggested to me that there is at least another way of ending the scheme for the story of *akrasia*. “X would have acted differently if he had not had a failure of nerve, if his vision of the good hadn’t, somehow, faded.” Developing this suggestion would require a full study. Such a study would, I think, show that failures of vision and nerve are best understood as involving both failures of character and of knowledge.

¹⁴ Donald Davidson, *op. cit.*, pp. 21–22.

¹⁵ I am indebted to Donald Davidson for his discussion of the problem of *akrasia*. Though he would almost certainly disagree with my diagnosis of the problem, much of my understanding of *akrasia*—and many formulations I use in describing the phenomena—comes from reading his paper. I have benefited from the comments of a number of people on an earlier draft of this paper. I am especially grateful to Gerasimos Santas and Gregory Vlastos for their suggestions and criticisms.

V. PURPOSE IN PAINTING AND ACTION

MARCUS B. HESTER

I

ARISTOTLE makes a cryptic remark which hints at some of the peculiarities of purposive moves in art. This remark is a good start for the specific interest of this paper—peculiarities of purpose in painting. Some of these peculiarities hold true for the arts in general. Aristotle says:

καὶ ἐν μὲν τέχνῃ ὁ ἐκὼν ἀμαρτάνων
αἰρετώτερος, περὶ δὲ φρόνησιν ἦττον,
ὥς περ καὶ περὶ τὰς ἀρετὰς.¹

τέχνη includes, of course, more than we mean by "art," but it is clear that he considered painting a τέχνη. W. D. Ross leaves the remark cryptic in his translation:

In art he who errs willingly is preferable, but in practical wisdom, as in the virtues, he is the reverse.²

Preferable to whom? To an agent who voluntarily does a bad act? Or to a painter who involuntarily erred? In his less literal translation Martin Ostwald emphasizes the latter:

In art a man who makes a mistake voluntarily is preferable to one who makes it involuntarily; but in practical wisdom, as in every virtue or excellence, such a man is less desirable.³

What it would mean to compare the artist to the ethical agent is not at all clear, and I believe the remark should be pressed in the direction suggested by Ostwald. If two painters, *A* and *B*, did identical bad paintings, *A* voluntarily and *B* involuntarily, *A* is a better artist than *B*. Given Aristotle's understanding of the artist as a master craftsman, we understand why he said this. An involuntary act calls for pity and pardon because the agent acts in ignorance of the particular circumstances. Though we will pardon ignorance of tools and clumsiness in action, we will not in art, for art is essentially

mastery of a craft. Technical skill and knowledge of the medium and tools are definitive of being an artist.

We attribute "wisdom" in the arts to the most precise and perfect masters of their skills: we attribute it to Phidias as a sculptor in marble and to Polycletus as a sculptor in bronze. In this sense we signify by "wisdom" nothing but excellence of art or craftsmanship.⁴

Further, we know from the *Poetics* that Aristotle held art to be, with some qualifications, a matter of imitating nature. Nature may be improved, but he did admire the skill needed to capture nature in this improved sense. To put Aristotle's point in another way: In art we expect the artifact to reflect the intention even if the intention was bad. In general, critics and philosophers have agreed that artistic purpose must pervade the artifact. John Ruskin states this belief in a very strong form:

For by a truly great inventor everything is invented; no atom of the work is unmodified by his mind.⁵

This paper will attempt to give some reasons for believing that the intention and realization in the artifact must coincide. In explaining this coincidence, we shall have to get into some more general questions about artistic purpose.

It is interesting and relevant to the question here that Aristotle calls medicine an art. The physician initiates a causal chain or chains leading, according to the laws of nature which he must understand, to the desired state, health in the patient. To what extent does the painter initiate causal chains? I shall argue that he does so only in a limited sense. The painter needs to know facts such as the effects of putting certain colors beside each other, and he needs to know the effects of color on distance in spatial composition. These causal chains are short. Painting is not so much a matter of getting

¹ Aristotle *Nicomachean Ethics*, vi. 5, 1140b23-24.

² *The Works of Aristotle*, tr. and ed. by W. D. Ross (London, Oxford University Press, 1949), IX, Bk. vi. 5, 1140b23-24.

³ *Nicomachean Ethics*, tr. by Martin Ostwald (New York, Library of Liberal Arts, 1962), p. 154. (I shall hereafter quote this translation.)

⁴ *Ibid.*, pp. 155-156.

⁵ John Ruskin, *Modern Painters* (New York, Merrill and Baker, no date), V, p. 182.

one thing to lead to another as it is a matter of getting a set of things to work together. The painter does not initiate a causal series which results in the painting. What he directly does on the canvas with his brush is causing only in the most strained sense. The painter's brush, as are tools in general, is an extension of his hand, and we do not cause our hands to move. In contrast, in deliberating about an ethical act, say opposing the war in Vietnam, one has to foresee very elaborate and complex causal chains and has to foresee the probable effects of the various possibilities. In painting, verbs like "foresee" have little employment. Instead, we use anticipatory verbs like "visualize," "imagine," and "picture." There is a difference in kind between the conceiving stage of an act and the conceiving stage of a painting, as we shall see in considering excuses a painter might give.

It is tempting to think that Aristotle slighted what we are at present vaguely calling the conception in painting.⁶ Is it not really just as bad for a painter to embody successfully a bad conception as to fail technically to realize his conception? Aristotle may not have raised this question because he assumes art is some form of imitation. Perhaps he should have said that a painter who involuntarily does a bad painting is no painter at all while a painter who voluntarily does a bad painting is a bad painter. Still it is not clear that being a bad painter is preferable to being no painter at all.

In summary, Aristotle clearly was aware that purpose in art is basically different from purpose in action; that voluntariness is of a different nature.

In the arts, excellence lies in the result itself, so that it is sufficient if it is of a certain kind. But in the case of the virtues an act is not performed justly or with self-control if the act itself is of a certain kind, but only if in addition the agent has certain characteristics as he performs it: first of all, he must know what he is doing; secondly, he must choose to act the way he does, and he must choose it for its own sake; and in the third place, the act must spring from a firm and unchangeable character. With the exception of knowing what one is about, these considerations do not enter into the mastery of the arts; for the mastery of the virtues, however, knowledge is of little or no importance, whereas the other two conditions count

not for a little but are all-decisive, since repeated acts of justice and self-control result in the possession of these virtues.⁷

(Aristotle plainly states that "voluntary" is a wider term than "choice."⁸ For our purposes, the most interesting voluntary act not chosen is the impulsive act. *Prima facie* there are many things that the painter does in painting which seem impulsive, many manipulations based on instinct and feeling. Such voluntary moves by the painter are not in actuality impulsive. There is a sort of "knowledge" or purpose behind them. I shall argue that this basic class of purposive moves in painting is not really assimilable to any action models of purpose.) We cannot evaluate an act without knowing whether or not it was voluntary, but in art we evaluate the product itself since, by definition, the craftsman has skill and knowledge sufficient to realize his intention. Ignorance and clumsiness are inexcusable in the craftsman. The artifact must reflect the intention.

II

Austin's essay on excuses was inspired by Aristotle, and it is helpful to use his distinctions to bring out some more peculiarities of artistic purpose. The study of excuses shows us that an action may go wrong in any stage from the appreciation or understanding of the situation to the execution of an intention. More of the peculiarities of artistic purpose will come out if we concentrate our attention on the executive stage, the stage where we might, as Austin put it, miff it.⁹ (I have already indicated that I consider errors in the conceiving stage as being at least as serious. However, these errors in conception would rarely be admitted to be errors. The painter would more often justify than excuse his conception.) The most relevant concepts in execution are mistakes, accidents, and inadvertences. By "mistake" Austin seems to mean a misidentification.¹⁰ To use his donkey example: Suppose you and I each have a donkey. I decide to shoot mine, and I see and kill the one I think to be mine. On closer inspection, it turns out to be yours. This is a mistake. Suppose again that I have drawn a bead on the donkey

⁶ I shall continue to call the early imagining, planning, and deciding stage of painting "the conceiving stage." No more precise a phrase can be used in view of the variety of ways paintings are conceived.

⁷ Ostwald, *op. cit.*, p. 39.

⁸ *Ibid.*, p. 58.

⁹ J. L. Austin, *Philosophical Papers* (Oxford, The Clarendon Press, 1961), p. 141.

¹⁰ *Ibid.*, p. 133n.

which really is mine. Before I can pull the trigger, he moves, and I kill yours by accident.¹¹ This accident is an error caused by an unforeseen happening and perhaps by my carelessness. Inadvertence is a different error still. Inadvertence is when in the process of doing act *A* (when I am not supervising in detail what I am doing), I cause event *B* when *B* is untoward.¹² In passing the butter, I inadvertently knock over the tea cup. These distinctions can be applied to things which happen while painting. I might want Prussian blue for this place but by mistake get ultramarine on my brush. I might add a color before a previous color is dry and get an accidental mixture which turns out unhappily. I might, as Cézanne did, inadvertently drop some splotches of water on the paper while reaching for the palette or while brushing too vigorously. We clearly can make mistakes, have accidents, and do things inadvertently in painting. However, the interesting question is if we would accept a painter's excuse in terms of any or all of these. Excuses for acts, if they are accepted, get us out of the fire into the frying pan. What would we do if a painter gave us such an excuse? I have a suspicion that Aristotle is right. Such an excuse would not get the artist out of the fire into the frying pan but out of the frying pan into the fire. What is inappropriate and unacceptable about this sort of excuse on the part of the painter?

The main reason excuses having to do with the execution of paintings are not acceptable is this: We will not accept the intention-accomplishment split or the intention-caused-to-happen split (inadvertence) that we accept in action. We must accept this split in action because the world is so various that we cannot anticipate factors beyond our control. We did not anticipate that our donkey was going to move suddenly while we were squeezing the trigger. Further, the range of skills (identifying and executing skills) we might be called on to use is so various that we can expect only an average performance from the average person. We might need to be an excellent swimmer or a fire fighter or a skilled driver. Since the range of things we might suddenly need to do is so wide, we can expect only a normal performance in any one of them. Exceptional skills can hardly be expected from the ordinary man. Therefore, we do accept an attempt-accomplishment split. We consider the attempt to save the child praiseworthy even if it

fails. And if the act is untoward, we consider the intention as well as what was caused to happen. There are some basic differences in the situation of painting which make such an attempt-accomplishment or intention-accomplishment split unacceptable. In the first place, we expect, as Aristotle did, that the painter is a master of his craft. The concept of what a master in painting is has certainly changed, but whatever the painter intends to do we expect him to do very well. The representational craft is not the only craft which can be mastered. The only factors standing between the painter's conception and its execution is his own skill and his understanding of the medium. He practices to develop his skills, and he studies his medium to know what it will do. The studio is an isolated environment. It is within the painter's power to eliminate the unforeseen circumstances. Despite our foresight, we cannot anticipate all the developments which start from our acts. There are unpredictable doings by other persons (or animals) and unforeseen causes of a physical sort.

In the second place, to approach the isolation of the studio from another angle, the painter has the possibility of seeing his painting complete and approving it before letting it into the world. We might roughly break down the process of painting into three stages: conception, execution, and survey of the finished work. They are not, of course, in actuality serialized in this neat way. Many conceptions arise out of experimenting with the medium. Further, painters usually step back and contemplate the work while executing it. The key point is that the painter is able to see the final result, to contemplate the whole. The actor cannot foresee the final result of his act. For a painter to send a painting with an excuse would be like an omniscient being trying to excuse his act. If he does not approve of the result, he need not sign it or let it out of the studio. When he signs it he is saying, I approve what I have done and this work comes up to my standard. The actor is not in a position to endorse his act, to give it his signature before it has had its effect. What we would give to be able to sign our acts! Because the studio demands the contemplation and approval of the work, the painter's excuse, if we allowed him one, would telescope into one similar to what Austin calls a mistake. Austin means by a mistake, it will be recalled, a misidentification. If we extend this concept to cover more general errors of judgment,

¹¹ *Ibid.*

¹² *Ibid.*, p. 140.

then it is suitable for the painting example. The artist may have made mistakes, had accidents, and done some things inadvertently. They all, however, fall under the final approval, and it is a mistake, not an accident, to approve of mistakes, accidents, or inadvertences (when there is no quality of the work which redeems them). To let an accident or inadvertence get out of the studio is a failure of judgment, not a failure of skill. The painter's final approval and signing of the work causes what we might call a process-product shift. What the painter finally approves is not a process but a painting. Ultimately the painter's point of view coincides with the spectator's. The intentionalist has not taken seriously enough this conversion. The process-product conversion renders the process of painting, including any unrealized intentions or alternatives not chosen, irrelevant.

It might be argued that the painter may commit something such as Austin describes as an accident in that he might not have foreseen the effect of his painting on the public. If one means by "effect" the moral effect, then we would be committing our painter to a Tolstoian view of art; and we would endanger the distinction between painting and propaganda. If one means the effect on the painter's popularity, the situation is still dubious. No great artist would have responded very much to a Harris poll. One does not have to be a purist to say that approving a painting is the natural end of a process not the beginning of an action. A painting is just not a cause in any normal sense, and again we expect no foresight from the painter. A basic problem in action is that acts will not fall into neat complete wholes as do paintings. Where an act begins and ends is a basic moral question. The painter does not cause a painting, nor is the painting a cause.

Finally, we would not accept a painter's excuse because he could always simply withdraw the work instead. We cannot withdraw actions, though we can make amends. Ethically and legally we try to compensate for things we have done which are untoward. We can withdraw intentions and promises, but not acts. The painter, on the other hand, can quite literally take back his painting. It is considered the height of insincerity if one makes an excuse for an act and does not attempt to right it. If acts could be withdrawn, we would

prefer that to any excuse. We would be incensed if the painter attached a note: "Dear viewer, please excuse the mistakes, accidents, and inadvertences I committed in the foreground."¹³

Austin at least mentions excuses having to do with the early stages of acts in contrast to those having to do with execution. There are stages of our acts involving "intelligence and planning, of decision and resolve, and so on."¹⁴ Then there is the very tricky early stage of appreciation, a stage where we may err through wrong emphasis even though we know all the relevant facts, errors due to "thoughtlessness, inconsiderateness, lack of imagination. . . ."¹⁵ This stage of an act promises to be fruitfully compared to what we have been loosely calling the conceiving stage of painting. There are many anticipatory or predictive verbs used to describe our deliberations about distant possible results of what we are thinking of doing. There are "foresee," "foreknow," "predict," "anticipate," "envisage," "conjecture" (just to name a few). "Foresee" will serve as a representative of this family of verbs. We notice immediately that a mistake of foresight is not really what happened in the mistake in Austin's example. (Nor is a mistake of foresight like the accident in that example.) To misidentify my donkey is hardly to have a failure of foresight. "Foresee" is just not the right word for immediate short-range causal consequences of my act. The prophet and the statesman need ability to foresee. I need no foresight to select my dinner, though I do need foresight to decide whether to support the Vietnam War or not.

"Foresee" has only a strained application to the painter, as I have already hinted. The painter does not need to foresee so much as he needs to visualize and imagine. In deliberating about which act we should do, we try to imagine what the resulting situations would be like. In thinking about a painting, the painter's anticipation is of a more specialized sort. He needs to know what the painting will look like or what effect some specific move will have. We shall need to analyze the special sensory verb "visualize." "Visualize," "imagine," "picture," or "image" are not in the predictive family of verbs but in the conceiving family. If I succeed in visualizing or picturing I need not succeed in knowing what is to happen.

¹³ There is one conceivable excuse a painter might make, namely, having a change of mind. But a change of mind is not the counterpart to a failure of foresight in action. A change of mind is just a change of mind, not a new awareness of consequences.

¹⁴ Austin, *op. cit.*, p. 141.

¹⁵ *Ibid.*, p. 142.

It is true that "visualize" may function in action when we try to picture what certain situations will be like. And we do often say "he was not as I pictured him" or "he was exactly like I pictured him," and these are actuality tied uses of "picture" or "visualize." "Visualize" may even be related to questions of causes, as when we try to visualize what will happen if we do thus and so. But we also allow "visualize" and "imagine" a non-reality function, and in painting this is the dominant sense. Painting conceptions have imaginative freedom and need not be connected with any actual possibilities. To conceive a painting is not to foresee anything. Even if a painter very concretely visualized a painting he later did, this would not be a case of prediction, for we do not predict what we intend to do. Imaginative visualizing is just not predictive.

However, I do not want to rest much weight on the specialized verb "visualize" or any other nonpredictive uses of "imagine" or "picture" since it may well be the case that some kinds of painting conceptions (plein-air landscapes, for example) do not require any such imaginative projections. Further, only introspection could determine whether or not painters visualize. Rather, I want to point out some factors in the nature of painting which in a more positive way show deep dissimilarities between purpose in acts and purpose in painting. If we look at some typical ways paintings are conceived, we discover, as the remarks about the uselessness of foresight in painting suggest, no counterpart of moral deliberation. Further, the executive stage, as the remarks on excuses suggest, is quite different from any action counterpart. A closer look at some well-known facts about artistic production will substantiate Aristotle's claim that purpose in painting is of a peculiar nature.

III

First, we must show that painting does not involve deliberation in any normal sense. It is not the purpose here to develop a profound analysis of deliberation in action, but rather to reach some undebatable (and thus dull) conclusions which will serve to contrast with some typical types of painting conception. As Aristotle noted, we deliberate about what we think are possible

courses of action. Taylor agrees that we deliberate about what we think is in our power, though he goes further to state some more questionable theses about the incompatibility of deliberation and determinism.

Deliberation, as I am conceiving it, is a process of active, purposeful thought, having as its aim or goal a decision to act, under circumstances in which more than one action is, or at least is believed to be, possible for him who deliberates.¹⁶

Deliberation is different from speculation and inference.¹⁷ In common-sense terms, we deliberate about the various alternatives which we think are in our power. We may even deliberate about whether or what are the chances that something is in our power, that we could do it if we tried. Some of this typical kind of pondering consists in making predictions about what will happen in a causal sense if we do each of the various alternatives. More importantly, we try to anticipate what other agents will do. However, deliberating is not entirely or even mainly a matter of prediction, and Aristotle, perhaps because of overemphasizing the physician's deliberation, interprets deliberation too much in terms of causal prediction. Kurt Baier agrees that deliberation is not mainly prediction when he relates deliberation to justification, rather than explanation.¹⁸ I also deliberate about how others will take my act, and this is hardly a form of causal prediction. Further, my act itself may be the most important thing (not any causal consequences of it). If I decide to oppose the constituted authority, my act itself, not its results, may be the heart of the matter, for it may be a symbol. We speak of such acts as being "matters of principle" even though we expect no causal results. Further, I think Aristotle was wrong in saying we do not deliberate about ends (though, admittedly, he has a special sense for the concept of ends). One of the most painful parts of deliberation may be coming to decisions about main and subordinate values, and the metaphor "weighing the pros and cons" clearly suggests that in deliberating a part of the deliberation may be coming to some conclusions about our hierarchy of values. Further, we deliberate about the "consequences" of our contemplated acts, and the word is revealing. "Consequences" covers not only what will happen and how others will take my act, but how it will

¹⁶ Richard Taylor, *Action and Purpose* (Englewood Cliffs, Prentice-Hall, Inc., 1966), p. 168.

¹⁷ *Ibid.*, p. 170.

¹⁸ Kurt Baier, *The Moral Point of View: A Rational Basis of Ethics* (Abridged ed.; New York, Random House, 1965), pp. 41-43.

resound in affecting me. I must be prepared to suffer or enjoy the consequences of my act, and parental and constituted authorities use this kind of consequence in our moral education. In conclusion, we deliberate about the consequences of what we take to be real alternatives open to us. We do this by predicting, guessing how others will interpret and react to our act, and guessing how the act will resound on us. We often must settle on some hierarchy of values. Deliberating is not, of course, the same thing as foreseeing, though good deliberations require foresight. Deliberation is no single kind of activity. In fact, it seems plausible to say that there are different kinds of excellence in action deliberations. The general and the statesman (in power politics) need excellence in predictive ability, in foresight, not in the sense of the physicist's prediction, but in the sense of predicting what will happen due to the manipulation of events by other agents. This ability is extremely complex, involving things like insight into character, understanding of potentialities or abilities of agents, sensing where key crossroads of events will be, etc. Churchill had this sort of deliberative excellence. The great morally sensitive person, like Socrates or Christ, needs an entirely different sort of deliberative excellence, one which requires a great feeling for priorities of values. An agent with tact needs a different sort of deliberative excellence still, a perceptivity about how others will feel about what he does. Thus there are different kinds of perceptivity required for the various types of things involved in deliberation in general. The most deliberative aspect of painting, conceiving a painting, seems to me very different from these senses of deliberation.

We must distinguish a conception of painting from a conception for a painting. For example, Cézanne's conception of painting was to do Poussin over again after nature. His conceptions for paintings vary with the individual case, being an expression of what he wanted to emphasize or bring out in this particular painting, given his general aesthetic point of view. Conceptions for paintings vary, of course, according to one's conception of painting, and thus there is no one process of conceiving of a painting. Painters may conceive of a painting in the sense of coming up with an original imaginative conception of a subject. Michelangelo conceived of "Creation" in this sense, and Delacroix, "Dante and Virgil." Painters conceive of paintings in the sense of planning overall composition. This sort of con-

ception may be worked out in actual sketches (as may the imaginative conception). A plein-air landscapist may conceive of a painting by selecting a subject, angle, season, and time of day. A painter may conceive of a subject in the sense of deciding what to emphasize in it. He may see a scene as a study in light and shadow or as a study in warm and cool colors. A painter may conceive of several styles of rendering the subject in a technical sense. Should the subject be treated in free calligraphy or tight washes? Should he emphasize planes or indistinct edges? Painters think in terms of their medium. These are some normal senses of conceiving of or planning a painting. There are only some limited strained similarities to deliberation in action. None of the above involve foresight or prediction of causal consequences. Nor do painters deliberate, as previously suggested, about the reactions to their paintings. In these senses of conceiving of paintings, there just is not a very real sense of entertaining alternatives and their consequences. The painter's motives and purposes are so bound up with medium considerations that his deliberation is of quite a peculiar sort. Handling qualities and technical questions are important considerations. A good way of putting it is that the painter considers what he can do to give the painting certain *qualities* (not what will be the results), and many of these qualities fall under his aesthetic. "If I do this, it will increase the luminosity" is a typical consideration, and medium motives are evident in it. "Merit" is another word which is especially used in the value considerations in productive disciplines. The most general ends then of painting are described in different terms from those in practical deliberation. Painters want to make paintings with qualities of originality, distinction, workmanship, etc. Such general ends really cannot be called "results" or "consequences." Things just do not initiate events in the world in the same ways as acts, and thus reflective production involves a distinctive type of deliberation. Perhaps "deliberation" is not even the right word for conscious and reflective production. The only exception is that painters do ponder aesthetic values or the direction they want their painting to go, and such reflections are analogous to deciding on relative values in moral deliberation. But even this sense of deliberation may well be *ex post facto* reflection on how one's painting is evolving. In general, our previous conclusions stand: Painting is not a matter of getting one thing to lead to another but a matter of getting a set of things to work together.

A painting is not the beginning of an action but the end of a process.

IV

But more of the peculiarities of purpose in painting come out if we examine the executive stage of painting. Here we see why there are no excuses in Austin's sense. Almost all critics devise some concept to deal with the kind of purpose involved in the painter's *response* to the developing painting. "Response" is a very good word here because we respond in a trained way in regions where rules do not apply or are not helpful. And response is a sort of purposive action which is appropriate to just this particular situation. Thus, even though there are some laws of visual phenomena, eventually these laws, such as that warm colors tend to advance, become academic rules. In order to provide that painters operate beyond such academic rules, critics usually build into their criticism some sort of response concept. Even the great academician Reynolds emphasizes the limitations of rules, and he provides response concepts in his concepts of taste and genius. Response is not impulsive behavior, for an impulse wells up within us. Thus we speak of being "seized by an impulse." In fact, very powerful desires, emotions, urges, or passions destroy the sensitivity and keen perceptual awareness necessary for response in our sense. A response suggests a keen awareness of the situation and action appropriate to just this situation. In some sports, such as tennis, quick and appropriate response is necessary, and it is acquired, to the degree that it can be acquired, by practicing the basic strokes so that one can concentrate on the subtleties of court position, etc. Practice frees one so that he can respond. Response in executing paintings has two aspects, and they are isolable only artificially. Responsive execution has a handling and a sensing aspect.

By handling I mean the qualities the artist's gesture imparts to the painting by actual brush strokes (or other instruments used). I have deliberately chosen a broad concept, one which can cover handling abilities as diverse as skill in the narrow representational sense or handling in the sense of free calligraphy. Handling can further include placing and gradation of light and dark.

Aristotle, of course, realized that the artist must have handling abilities, but he might not agree with the extension we have suggested to non-representational art. Handling may be praised by the purest painting type of critic as well. In fact, Fry's basic concept of sensibility is, in part, a handling concept. "Sensibility" designates the unique kind of purpose imparted to the painting by the artist's sensitive and responsive handling of his materials. It is shown "by the specific quality of his lines, by the relations of his tone and colour and by the handling of his paint."¹⁹ Thus, he seems to mean that it functions with regard to several elements of painting as they are manipulated in execution. In a similar manner, Ruskin praises the subtle modulations in the stroke of a great painter.

And, indeed, this delicacy is generally quite perceptible to those who know what the truth is, for strokes by Tintoret [*sic*] or Paul Veronese, which were done in an instant, and look to an ignorant spectator merely like a violent dash of loaded color, (and are, as such, imitated by blundering artists,) are, in fact, modulated by the brush and finger to that degree of delicacy that no single grain of the color could be taken from the touch without injury.²⁰

We need to note that activities involving handling abilities impart an indistinct periphery to the question of what was intended and what was not. The body may know things the mind does not, and even the painter may be puzzled by some unusually fine subtlety he executed. Skills are a backlog of knowledge in the painter's hand and arm. His hand and arm may do things transcending what he has done before, and he may be uncertain about whether he intended them or not. The wisdom of the trained body gives no little support to the inspiration concept of the artist. As Fry puts it:

And after all it does not much matter what the artist may have intended—what we feel is what came through from the artist's unconscious reactions, what transpires inevitably from the quality of his handling and the harmonies of his tone contrasts—and this belongs to the universal language of art which leaps across all divisions of space and time and puts us into direct contact with the artist's spirit.²¹

We also recall Ruskin's remark about the artist's

¹⁹ Roger Fry, *Last Lectures* (Boston, Beacon Press, 1962), p. 24. We notice here the artificiality of isolating handling from perceptive choosing, for sensibility is related to sensitivity, and sensitivity is shown in a painter's choices.

²⁰ Ruskin, *op. cit.*, III, 38.

²¹ Fry, *op. cit.*, p. 147.

purpose extending to every atom. One main reason philosophers and critics have almost universally believed that purpose in art extends to very small manipulations, that we can ask why of every atom, is that the wisdom of the trained body makes it impossible to say what was intended and what was not.

Artistic purposes which are implemented by various sorts of practicing are not the sorts of things for which one consciously intends something in this particular case. Even selecting a color depends on choosing sensitively rather than executing skillfully. One intends to develop capacities for handling subtleties, but this sort of intention is implemented in a different way (by practice) from implementing the results of deliberation, from acting on the basis of reasons. The important moral qualities of actions just do not require a high degree of executive skill. Instead of skill, we admire characteristics like persistence, making a serious attempt and deliberating well. We do not admire the style of execution in the way we admire the style of handling in Botticelli's line. Of course, there are other disciplines besides the arts where practices are used to develop knowing how skills, and in these there is precisely an admiration both of success and of style, as in a "beautiful serve." Games and sports often have quasiaesthetic dimensions.

I choose the phrase "sensitive choice" because "choice" can be used to cover nondeliberative types of voluntary acts as well as deliberative ones. Thus "choice" can cover artistic purposes of both the conscious deliberative sort and the spontaneous responding type. However, emphasis on sensitivity is really more important than emphasis on choice, for after all, in handling the painter chooses, and thus "choice" does not bring out the uniqueness of the aspect of spontaneous execution I have in mind. This aspect is more accurately characterized as a kind of sensitiveness or perceptiveness. "Sensitive" is, of course, an open-blank word, and we always want to know sensitive in what respect. The critic's aesthetic usually explains or suggests the kind of sensitiveness he looks for in paintings. The viewer is often said to need a nonproductive counterpart of the painter's sensitivity in order to appreciate what the painter has done. For Ruskin, Reynolds, Baudelaire, and Fry the painter and connoisseur both require respectively imagination and taste, and imagination and sensibility. Sensitivity is, of course, inherently a

concept which can sensibly be used of both productive execution and critical appreciation, while the handling aspect of response in painting has no nonproductive counterpart. It can reasonably be expected of us in our role as viewers of paintings to develop our sensitivities.

Ruskin tells an anecdote about William Hunt which gets at the heart of what is meant by "sensitive choice." Once while watching Hunt paint, Ruskin asked him why he put a certain color in a certain place. Hunt replied: "I don't know; I am just *aiming* at it."²² Aiming-at-it is a kind of purpose artists often mention when talking about executing paintings, and it is a sort of trained sensitive response to just this context in the developing painting. Of course, painters respond in more ways than in the handling of their color. The execution of subtle rhythms of calligraphy is another well-recorded type of response (as mentioned above).

Choosing in a responsive or sensitive way may contrast with all sorts of more pedestrian forms of knowing and purposing. It may contrast with a choice made on the basis of rules, for example, rules of composition. It may contrast with systems of construction, for example, systems of space construction. It may contrast with systems of selection as in systems of color harmonies. It may contrast with some pedestrian method of investigation, as in the laborious discovery of the principles of atmospheric effects. It may contrast with some elaborate symbolism of, for example, color. Sensitivities, as we would expect from their being sensitivities, are said to be innate possessions of the geniuses of painting. They can be developed but not really learned. Further, sensitivity, as did handling, makes the periphery of what the artist did on purpose and what not indistinct. Recognition of sensitivity makes it possible to extend the range of the artist's grounded choice.

Exactly how the spontaneous choice is claimed to operate depends on the critic's aesthetic. However, even though what responsive choosing means depends on an aesthetic point of view, different points of view share this general belief: That a painter in executing a painting responds on the basis of sensitivity to the pure elements of his craft in the sense of color, space, composition, and whatever else this list includes, itself not an undebatable topic. In other words, even non-purist critics like Reynolds, Ruskin, and Baudelaire believe that some of the painter's considera-

²² Ruskin, *op. cit.*, III, 86.

tions in responsively choosing in executing a painting involve intra-painting considerations of the developing context of the painting in itself. Purism simply exaggerates motives of painters which critics have always recognized.

It seems clear that there is no counterpart in action to material or medium sensitivity, and whether the actor and artist share some other kinds of sensitivity is beyond the scope of this paper. The uniqueness of material sensitivity is enough to establish that the painter's responsive execution is not really like such action counterparts as impulsive actions. Further, it must be emphasized that we are not here denying a creative dimension of morality. The phrase "sense of justice" needs investigating. If we imagined an Oriental tyrant who, because of the fairness of his rulings in disputes, was said to have a sense of justice, he would have qualities like a feeling for rights and priorities of rights, an ability to transcend statute law (if any exists), and a feeling for just compromises. Certainly there is a creative dimension of morality, but I argue that phrases suggesting such creativity as "a sense of justice" do not contrast with the same kind of thing as the various kinds of sensitivity in painting. A sense of justice mainly contrasts with slavish obedience to statute law. Further, we recall the different kinds of perceptivity required by the general, statesman, and great moral leader. Sensitivity in painting contrasts with very different things such as systems of space construction, systems of coloring, non-natural or conventional symbolism, and pedestrian means of investigation. Thus, it seems to me plausible to say that the sensitivities required for creative art and creative morality are different in kind.

V

As mentioned earlier, it was only with artificiality and for purposes of analysis that we distinguished the aspects of handling and sensitively choosing in executing paintings. Now we need to recombine them for some general conclusions, and Fry's concept of sensibility well expresses response in handling and choosing during the execution of paintings. It must be emphasized, however, that the word "sensibility" is hereafter used in a broader than puristic sense, for I shall use it to

cover the kind of purpose in execution meant by such diverse critics as Fry, Ruskin, Reynolds, and Baudelaire. We also recall that all critics have recognized artistic motives in the handling of the pure elements of painting, though in nonpuristic critics these pure painting means have been handled and sensitively selected in order to further some higher ends of painting.

There are an extremely large number of idioms from critics of widely different points of view which suggest some kind of sensibility in this new sense. I have in mind ones such as "has a keen sense of —," "shows real feeling for —," or "intuitively grasped the principles of —." We need not belabor the point with numerous examples since they are so easy to find. Two will suffice. Baudelaire asks what posterity will say of Delacroix, and he answers:

Like us, Posterity will say that he was an unique meeting-place of the most astonishing faculties; that like Rembrandt he had a sense of intimacy and a profoundly magical quality, like Rubens and Lebrun a feeling for decoration and combination, like Veronese an enchanted sense of colour, etc.; but that he also had a quality all his own, a quality indefinable but itself defining the melancholy and the passion of his age. . . .²³

Fry's whole *Last Lectures* are a demonstration of the presence or absence of sensibility (in his sense) in works of art. He does not want to attribute a consciously reflective artistic understanding to the works of primitive preliterate peoples whose work he loves and highly evaluates. Instead he talks of their sensibility, instinct, or feeling. Of a Scythian bronze pole ornament of unnamed origin, he states:

Very few people have ever had so profound a feeling for vital rhythms that they could reduce all the complexity of the forms of a living being to so severely simplified a statement.²⁴

We see in these examples the extreme variety of things sensibility has been used to cover, qualities as diverse as a sense of color to a sense of vitality. Thus the justification of our use of sensibility to cover a wide range of responsive handling and sensitively selecting aspects of painting. What exactly do idioms like these suggest?

First, when critics speak of an artist's feeling, intuition, or sensibility, they are not attributing

²³ Charles Baudelaire, *Art in Paris 1845-1862: Salons and Other Exhibitions*, tr. and ed. by Jonathan Mayne (London, Phaidon Press, 1965), p. 143.

²⁴ Fry, *op. cit.*, p. 124.

any episodes to his mental life. To say that the Scythians had a feeling for vital rhythms is not to say that any feelings, in the episode sense, occurred in their mental life. "Feeling" as used by the critic is closer to the perceiving use of feeling than to the mental occurrence use of feeling. The artist is claimed to have a feeling for his material, not to be experiencing any sensations. Idioms like "felt," "intuited," or "sensed" are praise terms having to do with some purported sensibility shown by the actual painting. The critic has a special job in making such claims stick, for he must make explicit the handling qualities and kind of sensitivity involved in the painter's execution, and these motives are notoriously difficult to state.²⁵ These idioms do have this much to do with the process of painting: If, in general, artists did not make purposive moves (though not necessarily purposive in the sense of entertaining alternatives or verbalizing intentions) based on sensibility, it would be misleading to use these idioms. These idioms are committal to a general belief in the existence of motives based on sensibility in the process of painting. They are not committal to the occurrence of a special awareness having to do with the part of the painting in question. Even the painter's conscious awareness is no reliable guide, for sensibility is not a reflective but an intuitive response. The painter may well have no special feeling when he follows his sensibility. The critic covers himself from such attribution of episodes in that these idioms usually operate under theories developed by the critic about sensitive faculties. A faculty is not an occurrence. Sensibility for Fry, imagination for Baudelaire and Ruskin, and taste for Reynolds are faculties, not events. When theories of faculties are overthrown it is not because of reports painters make about what occurred in their mental life while they were painting. It is when the kind of disposition specified in the critic's theory of sensibility is no longer relevant to the production and evaluation of paintings. As mentioned earlier, the whole process of conceiving of and executing paintings changes radically. It is this change which overthrows critical theories about special sensibilities. Few

critics of painting have really committed what is known as the intentional fallacy because few of them really speculate about what the painter thought while painting. To conclude, idioms like "had a sense of —" or "had great feeling for —" suggest a disposition to perform excellently in the specified respect, and such idioms usually operate under any theory of special faculties the critic may hold. They suggest sensibility as defined by the critic.

It is philosophically significant that this sort of purposive response called "sensibility," which sees such use in critical explanations, rarely, if ever, occurs when we explain our actions. As noted above, response does not mean impulsive reaction, and perhaps impulse is the nearest (not very near) action counterpart to painting response. However, here it is not claimed that we do not speak of response in action. We do. It is that we do not speak of response in the sensibility sense. Responding sensitively to the aesthetic dimensions of the environment of our act is just not important in accomplishing a deed. Material sensitivity and bodily response are important in the dance and drama, but these are art, not actions. Both the quality of handling and the sensitivity of selecting are essential in painting. In action, small considerations of the style of my act are unimportant. What counts is not the style of my walk (as in style of handling), but my getting there or trying. Nor in my action do I choose sensitively with regard to the aesthetic qualities of any materials which might happen to be involved in my act. We are thought perfectly complete moral agents even though we may not possess in any high degree any special faculties such as taste, sensibility, or imagination in the sense that these terms are used by the critics cited.

How is the conceiving stage of painting related to responsive execution? By now it is obvious that there is no one answer to this question. We have seen that there is a plurality of basic conceptions of painting, and painting is not really a single activity but a plurality of activities, some of which exist only in peculiar types of conceptions of painting. The critic or painter makes statements of

²⁵ For Ruskin to cite Hunt's response is to give no explanation, and Ruskin himself must provide the reason why Hunt did what he did. To show that it was a grounded choice the critic has to make explicit the wisdom of the choice which the artist would only express as "I am just aiming at it." The artist may answer our why question only in terms of the choice feeling right. (Hunt's answer does not preclude reason giving just as "I hit him because I was angry" does not preclude "I hit him because he insulted my wife." Sometimes we do admit we had no reason for our act; we make no attempt to justify it. And often we explain these acts [but do not justify them] in terms of a powerful emotion, passion, desire, or impulse.) The critic must attempt to verbalize the reason. But ultimately we must *see why* the painter did what he did. The choice must speak to our (non-productive) sensibility. Talking can only do so much.

what the activity of painting includes in such aesthetic formulas as "the true end of painting is —." The plurality of activities comes out if we briefly consider the conceptions of painting held by Ruskin, Baudelaire, and Fry, and they are respectively that the painter's purpose is to capture truth to Nature, to express a dominant personality and to manipulate the pure plastic elements of painting. Under these major heads there are very different subactivities and very different sub-subactivities, covering purposes of painting from abstract conceptions of what painting should be to manipulating actual paint in the final executive stage. For example, capturing truth to Nature might include purposive subactivities such as observing, copying in an exact sense, matching the colors of the subject, capturing just the right atmospheric effect, and expressing the deeper truth of the scene. There are systems and practices which will help develop each sort of subactivity. There might be some special kind of sensibility required (Ruskin calls it "imagination"). Some overlapping, but also some different activities, would be called for in Baudelaire's form of the expression theory. According to Baudelaire, the artist does not so much capture as find elements in Nature, like words in a dictionary, and he imposes his temperament on the work. We would not expect the same sorts of subactivities here, nor the same kinds of observing or practicing or development of systems, though, as mentioned, there may well be overlaps. Baudelaire believes that the hand must be ready to be the slave of the brain, and some of the same practices necessary to develop capturing skills might aid this obedience. Fry discusses still different subactivities under the head of manipulating the pure plastic elements of painting. Now there is no matching or expressing. Instead there are activities like space construction, sensing the dimensions of space, exploiting color harmonies, and handling color in a plastic sense. Here there are still different observations, practices, and systems. In each of the critics, painting is conceived of as some basic activity presiding over a set of subactivities and sub-subactivities. Thus, different conceptions of painting call for different handling abilities and even different kinds of sensitivity. Very different practices are involved in developing the handling abilities, and the sensitivities are developed in different sorts of exercises. Painters conceive of painting in ways as diverse as the painter of imaginative, religious,

or historical subjects to the plein-air landscapist. Some contemporary painters, we might call them "medium painters," really begin with no conception, but let the idea develop out of suggestions the medium itself makes. In short, conceptions of painting are extremely diverse in type, and we must emphasize that conceptions of painting are not divorced from actual execution motives or from intermediate types of considerations such as methods of practice. The painter thinks in terms of his medium. Thus, his conceptions of paintings are thoroughly pervaded with considerations of handling and sensitivity. The painter's sensibility does not suddenly begin to function when he takes up his brush. Some painters even claim to see in terms of their medium. Thus, conceptions of paintings have all the peculiarities and uniquenesses emphasized about sensibility.

In view of this diversity, how can we make any significant comparisons to action purposes? We certainly cannot say that it is inconceivable that painting could involve anything like deliberation in an action sense. No doubt some medieval painters considered painting to be a visual sermon, and thus the painting was intended to have a certain moral effect, and thus the whole dimension of excuses would have been relevant. In a very different form of life from ours, paintings could be vehicles for moral reform. Further, we could conceive of a form of life where the main moral dimensions of acts were aesthetic, where the style of the act and even some form of material sensitivity was its good. But these conceivable forms of life would be very different from our forms of life now.

Now perhaps it will be helpful to requote Aristotle on differences between the moral agent and the artist. Of the agent:

... first of all, he must know what he is doing; secondly, he must choose to act the way he does, and he must choose it for its own sake; and in the third place, the act must spring from a firm and unchangeable character. With the exception of knowing what one is about, these considerations do not enter into the mastery of the arts.²⁶

Now, we can add that the artist's knowing what he is about is of a different sort from the agent's. In the conceiving stage of painting, the artist plans what he will do, but this planning is not like moral deliberation. Further, in the executive stage, the artist's knowing what he is about includes the unique dimensions of handling abilities (de-

²⁶ Ostwald, *op. cit.*, p. 39.

veloped by various sorts of practice) and special types of sensitivity. Handling abilities are knowledge how, and neither they nor sensitivity need entail any special conscious events. What the painter thinks to himself during execution is quite irrelevant to handling abilities and sensitivity, though both handling abilities and sensitivity are goals the artist deliberately pursues. Purpose in painting is more like drawing on a reservoir of skills and sensitivity than it is like weighing up and settling on a course of action.

To sum up our results: Large numbers of critics and philosophers have believed purpose radically pervades the painting. This belief is supported by the oddity of excuses in painting. A painter does not cause a painting, nor is the painting itself a cause, and thus the painter does not need foresight. The studio is an isolated environment. The painter can eliminate the unforeseen by knowing his medium, and he can practice to control his

moves. He can contemplate his result as God can foreknow perfectly. A painter does not deliberate in the sense of predicting the ramifications of his painting, attempting to guess the reactions of others or calculating how it will resound in its consequences on him. The painting is the end of a process, not the beginning of an action. Some positive facts about artistic production explain why the critic believes purpose pervades the painting and why purpose in painting is distinctive. Both handling abilities and sensitivity make the periphery of what was purposed and what not indistinct. And style of handling and kinds of material sensitivity are not relevant to the moral qualities of our acts. Uniqueness of artistic purpose is shown by the necessary appendages of systems, methods of practice, experiments and investigations, all designed to develop handling qualities and sensitivity.

Received November 5, 1968

Wake Forest University

VI. THE BURDEN OF PROOF

ROBERT BROWN

IN philosophical discussion the course of argument is sometimes impeded by each side claiming that the *onus probandi* now lies upon the other. Each side does so on the ground that its views, at this point, are *prima facie* more plausible than those of its opponents. Each side appeals to the other to take up the burden of proof and thus permit the argument to proceed. Usually these appeals are met by one of the participants and the discussion continues. But why should one side respond rather than the other? Given the conflicting claims, it is worth asking whether there is a legitimate way in which the issue can be decided. Are there any conditions under which *onus probandi* can be appropriately assigned in a philosophical debate? Is laying the burden of proof on someone a defensible procedure or merely a labor-saving device?

I

Now it would be natural to suggest that certain philosophical statements (or propositions) resemble legal presumptions to this extent: both represent conclusions drawn from premisses, explicit or implicit, concerning matters of common experience. The law presumes sanity in testamentary cases because, for one, people who make wills are usually sane at the time. This is, or is believed to be, the common situation. In philosophy there are general propositions which are either summaries of, or inferences from, certain beliefs of common experience. In both philosophy and the law, then, the presumptive character of certain claims is derived from propositions accepted by the world at large as truths of daily life. Therefore, someone might conclude, the example of the law indicates where we must look for at least some philosophical presumptions. They are to be found, for example, in the statements defended by G. E. Moore as general truths that are entailed by certain more specific truths of everyday experience. Instances of such general truths are: "Time is real"; "Space is real";

"There are material things"; "Someone sometimes perceives a material thing"; "Some material things exist unperceived"; "There are other minds"; "Some empirical statements are not hypotheses."

Moore's own view was that a statement like "There are external objects" is both empirical and true. He thought that some philosophers had mistakenly believed such a statement to be self-contradictory; and that from this false non-empirical claim these philosophers invalidly inferred the truth of the empirical statement "There are no external objects." Thus, said Moore, these philosophers were making both a false non-empirical statement and a false empirical statement. It was against the latter component that Moore's disproofs were directed: for example, the disproof of the statement "There are no material things" by Moore holding up one hand and saying "This hand is a material thing; therefore there is at least one material thing." In uttering these words and raising his hand Moore took himself to be producing a counterexample to a false empirical statement and also to be uttering a true empirical one.¹

On Moore's interpretation, then, the statements he was attacking are denials of some general truths that are entailed by truths of common experience. These denials are paradoxical not because they are verbal recommendations which go against customary practice, but because they are denials of such ordinary, reasonable beliefs as that temporal relations, material things, and other minds, really do exist. So someone who agreed with Moore's diagnosis might be led to think that we have in these general beliefs—closely allied to Reid's First Principles of Contingent Truths—something like a set of *onus-assigning* statements.

But there are two obvious drawbacks when we try to use such phrases as "in accord with common experience" and "general beliefs tacitly accepted" as characterizations of such a set. First, these phrases do not isolate the class of philosophically interesting cases. We need to distinguish between:

¹ See Moore's "Reply" in P. A. Schlipp (ed.), *The Philosophy of G. E. Moore* (Evanston, 1942), pp. 668-674.

"There are other minds" and "There are other inhabited planets"; "Time is real" and "Telepathy is real"; "Some material things exist unperceived" and "Some material things exist unperturbed." In each pair the second member is a straightforward empirical statement. It is open to refutation by counter-example and the nature of this counter-example is obvious in each case. In contrast, the first member of each pair is not, despite some of Moore's remarks, quite so straightforward in character. Each of them raises questions about its meaning; each of them can be, and has been, interpreted as analytic, as self-contradictory, as empirical; each requires us to think about the defining properties of its subject and about the propriety of ascribing a particular predicate to it.

Secondly, since these questions of interpretation are typical of, and help to define, the philosophically interesting cases, it is often a matter of controversy whether a particular philosophical belief is entailed by the beliefs of common experience. Thus in the last 40 years a great many lecturers have held up their right hands and (making a certain gesture with their left hands) said "Here is one hand, and this hand is a material thing." Have they thereby proven, as they thought, that there is at least one material thing? It has at least been debated. And where it has been, the denial of the entailment has had as its consequence the rejection, in turn, of "There are material things" as an *onus* assigning statement—rejected to the extent that the power to ascribe *onus* is supported only by the attempted proof. Of course the denial of the entailment must be supported by argument; it must be shown why entailment does not hold, and some indication is required of what *would* entail "This hand is a material thing." But to admit this much is to say that in the situation described *onus* falls on the person who rejects the proof given. Why should this be so?

We need to remind ourselves of a simple truth. When we read lists of what are said to be statements of common sense, we are reading lists of philosophically familiar sentences. And when we read papers like Moore's "Proof of an External World," we are to understand that his utterance of the sentences "Here is one hand" and "Here is another" is accompanied by the holding up of his two hands—and "a certain gesture" with each. The force of the argument consists not merely in the uttering or saying of certain sentences, but in saying them under certain conditions, conditions under which they are said to be both true state-

ments and statements of common sense. However, these conditions are not *described* by the lecture-room performer; they are merely exemplified in his performance of holding up his hands and making certain gestures. He gives us evidence, or reasons, for the truth of his claim—though he does not say what his evidence is—in a way that he would not have if he had held up some objects that resembled hands only vaguely. We rely on the truth of such a claim only when we believe it to be made under certain conditions. Yet since a description of these conditions cannot be written into the sentences which express these statements, it is easy to forget that sentences like "Here is one hand" are not always used to express true statements of common sense, that such sentences are sometimes used to make false statements or to make statements whose truth is in dispute. And forgetting this, it is even easier to forget that when the truth of the statement is in dispute, and the evidence is not clinching, the burden of proof must fall on the side with the weaker evidence.

Thus in the case of someone who presents Moore's proof, the weaker side must consist of those who, after listening to the proof, reject it on the grounds that the speaker has not proven that he has held up his hands. For by doing such things as listening to him with understanding, and by identifying him as the speaker, the audience have acknowledged both his presence and his attempts to give them reason to think his claim true. If they then wish to reject that reason, the *onus* is on them to produce better reasons which tell against that of the speaker.

Yet perhaps we are still interpreting our problem too abstractly. If there is no *set* of propositions commonly accepted as *onus*-assigning by all philosophers, there may still be some propositions or procedural principles that are commonly accepted by philosophers who find themselves at certain stages of particular arguments, or at typical points within specific kinds of arguments. Thus it may be a mistake to search for general propositions by which *onus* can be assigned. Suppose that the legitimacy of *onus*-assignment depends on less general considerations: for example, on each side agreeing that the premisses of one side give rise to a problem not produced by the premisses of its opponent. It may seem clear that if we examine a number of actual cases in which burden of proof is ascribed, much of our difficulty will solve itself. Consider then, the following three examples.

II

First is a characteristic appeal to a procedural maxim.

My thesis is that since the *onus probandi* lies on the philosopher who denies univocity [concerning "location" as applied to physical bodies and to sensations respectively], and since, according to what follows below the relevant equivocality is unproved, univocity ought to be assumed . . . it seems to me to be sound procedure to assume that a term is not ambiguous until some positive reason is given for believing that it is. But it also appears that sound procedure should restrict the nature of such reason. . . . Hence it seems that in the specific onus-principle that I enunciate above, "some positive reason" should be understood as "some positive reason based on suitable semantic considerations."²

This view obviously requires two amendments: it must refer to ambiguity of sense, not of reference; and it must refer to ambiguity of use on a particular occasion, not to the ambiguity (or diversity) of sense that many, or perhaps most, words display. Insofar as ambiguity raises a question here, it is the question "Which sense of the given word is intended on this occasion?" Sound procedure, then, is said to consist in assuming that a term is not being ambiguously used, on some given occasion, "until some positive reason is given for believing that it is."

But how can this assumption be sound procedure? If the view is supposed to apply specifically to philosophical arguments, then it seems likely to perpetuate the traditional muddles generated by its bland acceptance. Any reader of Locke, Berkeley, and Hume would be better advised to suspect ambiguity whenever he came across such pet words as "idea," "impression," "matter," "substance," "sensation," and "exists." The "positive reason" for doing so is that these and many other terms were in fact ambiguously used by the three authors, and that bad arguments were produced in consequence. Philosophers elsewhere are no different. Because ambiguity is such a common source of confusion in philosophy, the only sound rule is to suspect ambiguous use on all possible occasions.

To this it will immediately be replied that the phrase "on all possible occasions" encapsulates the problem at issue. Is every expression to be suspect? The only genuine question, surely, is

whether we can identify, beforehand, a class, or at least a collection, of likely suspects within certain types of arguments. If we cannot, then we have a defensible reason for claiming that the burden of proof lies on the philosopher who rejects the assumption of univocity. For it is clearly absurd to question the univocity of each term, on every occasion of its use, merely because some terms are sometimes used ambiguously.

The answer to this objection is twofold. First, not every expression is suspect. Given the circumstances of a special debate, we can usually identify, beforehand, a set of terms that need to be treated with caution. That is, we can usually provide some positive reason for assuming the likelihood of ambiguous use in a given context. Anyone who read Jaspers without suspecting, beforehand, his use of "communication" would be as foolhardy as a current reader who did not scrutinize the term "use" in Oxford philosophy of a decade ago. Shrewdness in philosophers consists, to some extent, in knowing what sorts of terms to suspect of ambiguity on certain sorts of occasions. It is this which makes it sound procedure to assume ambiguity rather than univocity. The need for either assumption must arise within limiting conditions, and in philosophy the dangers of ambiguity are of more immediate concern than the truism that, in general, most words are not ambiguously employed.

Secondly, if the assumption of univocity is supposed to apply not merely to actual philosophical debates but to any use of any term, then the assumption is empty. For to assume that the intended sense of a term is clear is simply to assume that we understand its intended sense. But this is an assumption we must make to carry on ordinary communication in daily life. We can hardly assume that we do not understand which sense is being used of each term that we employ—unless given "some positive reason based on suitable semantic considerations" for believing that we do. We could not then sensibly speak to each other unless given some positive semantic reason for believing, as we uttered each word, that we understood it. But if the reason is supplied either simultaneously with, or after, the utterance of the term, it comes too late for determining the choice of that term rather than some other term equally doubtful. And if the assumption of ambiguity is to operate, we cannot give a reason for believing,

² M. C. Bradley, "Two Arguments Against the Identity Thesis" in R. Brown and C. D. Rollins (eds.), *Contemporary Philosophy in Australia* (London, 1969).

prior to our employment of the term, that we understand in which sense it is being used.

Thus the assumption of univocality cannot, in its general form, be a procedural rule for dealing with troublesome cases. It merely describes one of the necessary conditions of linguistic communication. Since no alternative procedure is possible, it is pointless to recommend its adoption. On the other hand, in its specific form the assumption is obviously unsound, and has less claim to be adopted by philosophers than has its contrary. Of course, this is not to say that on some given occasion the *onus* of proof does not lie on the philosopher who says that a particular term is ambiguous. For in such a case we may be able to give good reasons for assigning the burden of proof. But the assumption of univocality will not be one of those reasons.

III

My second example is drawn from a paper entitled "Grünbaum on the Duhemian Argument" by Dr. L. Laudan.³

In the paper Laudan suggests that, taking '*H*' for "hypothesis," '*O*' for "observation statement," and '*A*' for "nontrivial auxiliary assumptions," Duhem held the following thesis: "In the absence of a proof that no appropriate hypothesis saver exists (i.e., unless we prove that $\sim(\exists A')(H+A' \rightarrow \sim O)$, then $\sim O$ is not a conclusive refutation of *H*, even if $H+A' \rightarrow O$." Laudan goes on to defend this thesis:

To continue to maintain *H* in the face of $\sim O$ is not necessarily to assert that a suitable *A'* exists, but simply to allow for the possibility that *H* may still be compatible with $\sim O$, given some suitable *A'*. The *onus probandi* is not, as Grünbaum supposes, on the scientist who refuses to call a refuted hypothesis false to show that his hypothesis can be saved by some suitable *A'*. Rather the burden of proof is on those who deny *H* to show that there does not exist an *A'* which would make *H* compatible with $\sim O$. Schematically, the scientist who claims to have falsified an hypothesis, *H*, must prove that $\sim(\exists A')(H+A' \rightarrow \sim O)$. Unless such a proof is forthcoming, a scientist is logically justified in seeking some sort of rapprochement between his hypothesis and the uncooperative data.⁴

Clearly, the thesis comes from, and helps to support, a most conservative policy for the testing of theories. The policy is conservative because it encourages the assumption that an *H*-saving *A'* exists until proven otherwise. And this assumption amounts to protecting *H* from $\sim O$ as long as (or longer than?) is decently possible: that is, amounts to conserving hypotheses from refutation by observations. The burden of proof is thrown upon those who think that a particular hypothesis has been refuted by observation; those who think it can be saved by an auxiliary assumption are not asked to produce a suitable *A'*. They are allowed to cling to *H* until their opponents prove that such an *A'* does not exist.

But how is this to be proven? Certainly not by direct observation, since even the most exhaustive search of the universe logically could not falsify the claim that a saving object or situation exists which is describable in *A'*. And if *A'* is supposed to assert that a saving law, otherwise unspecified, will be found in the future, *A'* remains unfalsifiable. So the refutation of *H* requires at the least, according to Laudan's Duhemian thesis, the advancement of another hypothesis, *H*₁, which is better supported than *H*; and the truth of *H*₁ must either entail or make highly likely the truth of $\sim(\exists A')([H+A'] \rightarrow \sim O)$. The hypothesis *H*₁ must weaken, to some high degree, the likelihood that a suitable *A'* can be found.

By why should supporters of *H* accept *H*₁? They can always say, as P. K. Feyerabend does:

... that the fate of many theories strongly depends upon the belief of their defenders that it will be possible, at some future time, to incorporate into them all the apparently refuting instances. Take, for example, the early belief that both the fixed stars and the planets obey the same laws of circular motion. The obvious irregularity of the motion of the planets was strong refuting evidence against this belief. Yet it was hoped that this apparently refuting evidence would somehow be explainable on the basis of the idea of circular motion, and the attempt to do so finally turned out to be successful. ... Hence it is not much of an exaggeration if we say that the progress of science frequently depends on the belief in the validity of the conclusion of Duhem's argument.⁵

³ *Philosophy of Science*, vol. 32, (1965), pp. 295-299.

⁴ *Ibid.*, p. 298.

⁵ Comments on Adolf Grünbaum's "Law and Convention in Physical Theory" in H. Feigl and G. Maxwell (eds.), *Current Issues in the Philosophy of Science* (New York, 1961), p. 158.

With this in mind, supporters of *H* can argue, as Feyerabend puts it, that "as long as different points of view are encouraged it is difficult to see how any refutation could be decisive in the sense that an alternative account which retains *H* but is not *ad hoc*, could be ruled out forever."⁶ Yet this is precisely what Laudan asks opponents of *H* to do. They are to prove that the refutation of *H* is decisive by proving that there does not exist a suitable alternative account which retains *H*. And here "does not exist" has the force "can be ruled out forever," since otherwise "does not exist" will mean "does not exist now"—a meaning which will not allow the refutation to be decisive.

Obviously there can be no burden of proof laid on a person to show what cannot be proved. So Laudan must be mistaken in ever asking someone to prove that no suitable *A'* exists. But is he also, and always, mistaken in placing an *onus* on those who reject *H*—given the acceptance of the Duhemian thesis by both sides? Surely not; for if *H* is a well entrenched and important set of theories, then the consequences of abandoning *H* will be so far reaching that its opponents will need more than some stray, apparently refuting, observations to overturn it. Critics will either be ignored or they will be told to produce a better theory. The latter demand will be justified by an estimate of the scientific gains and losses attaching to the abandonment of *H*. And since the Duhemian thesis is accepted by the critics, they will also have to accept the *onus* of proof unless their estimate of gains and losses is very different from that of the supporters of *H*. If it is, the critics will then be arguing that *H* is not well entrenched, that the consequences of abandoning *H* are not far reaching, and that, therefore, *onus* falls on the person who clings to *H* despite observations which seem to refute it. Often, disagreements of this kind about a particular application of Duhem's thesis can be, and are, settled by further examination of the supposed consequences and of the criteria of their importance.

Nevertheless, it is an empirical truth that while scientists often look for and find a suitable *A'* for saving *H*, they quite commonly do not. Instead, they turn to the development of a new hypothesis, and are not deterred by any belief to the effect that it is always empirically possible that an *H*-saving assumption exists. They are not deterred because the thesis is useful only as a means of defense against critics of *H*. In itself the thesis

provides no *specific* reasons for retaining *H*. But if there do happen to be specific reasons, then the thesis, expressing as it does a procedural rule, offers a more general reason why the attempt to repair *H* cannot be forbidden. Hence, we can disagree as to whether there are good specific reasons for retaining *H* even though we agree on the Duhemian thesis.

In this situation we shall agree on the ascription of *onus* only if we can come to agree on the soundness of those specific reasons. If we cannot, then the thesis will not help us. It does not enjoin us always to seek an *H*-saving assumption. What reason could we have for doing that? It merely states that there is no *general* reason forbidding us to do so. And to the extent that some past successes provide a general reason for future searches, the thesis does give us such a reason. It is not, of course, a strong reason when placed against more specific competitors. True, the progress of science may often depend on the belief that a suitable additional assumption can be found for *H*. But at least equally often, that progress depends on the contrary belief—as the history of science makes only too clear. Given that we accept Duhem's thesis, we can use it to ascribe *onus*; but only when more specific, and overriding, considerations are absent.

IV

Now it may seem that the two examples so far discussed are not the strongest cases available. Many philosophers will think that obviously stronger cases arise in connection with the principle of verification and the principle of parsimony (Occam's Razor). Thus it may be said that some version of both is a necessary condition of rational inquiry; and that, therefore, the *onus* of proof lies on anyone who claims that an entity exists which is neither observed nor observable. Similarly, it will be urged, the burden of proof lies on anyone who claims that two (or more) entities exist when only one is required for the explanation. Since for our purposes here the two cases present much the same problem, we can make a discussion of the latter stand for both.

Appeals to the principle of parsimony have played an important part in recent debates on the contingent identity of properties. For example, Max Deutscher gives some instances "where a thing's being *F* simply is its being *G*, and yet

⁶ *Ibid.*, p. 156.

so-and-so is *F* neither entails nor is entailed by so-and-so is *G*." He then continues:

But since a difference in meaning is not in itself a sufficient reason to claim distinctness in properties, it is up to the critic of physicalism to prove non-identity. (This last remark supposes that he does share the view that in science and philosophy we must not multiply entities beyond the necessity of accounting for agreed facts.) . . . So long as the physicalists' account of a mental state as that which has certain sorts of causes and which has certain sorts of effects is adequate to identify those brain states which exist when and only when the subject is in that mental state, it is up to the dualist to find something specifically non-physical about the mental state. If he cannot, he has no reason to deny the identification, and every reason to accept it.⁷

Deutscher says that since predicate-synonymy is not a necessary condition of property-identity, the burden of proof lies on the antiphysicalist to show the nonidentity of brain states and mental states. But this assumes that the argument from nonsynonymy of predicates is, in the debate which Deutscher carries on, the only objection put forward by his opponents. It is not, of course; even the objection from nonidentity of properties is often stated independently of that from non-synonymy of predicates. In addition, there are the many other objections which Deutscher himself mentions and rejects: for example, that mental states can be incorrigibly reported by the subject, that their descriptions are logically connected to descriptions of behavior, that mental states are intentional. Nor can it be claimed that *onus* is being ascribed here to a person who, at this stage of the argument, has already abandoned other objections. In Deutscher's paper they follow, rather than precede, the objection from non-synonymy of predicates. Moreover, still earlier in his paper, the author suggested that the failure to understand the identity theory could only be corrected by its supporters undertaking the burden of explaining how brain states and mental states could be identical.⁸ Thus according to Deutscher's own program, his ascription of *onus* ought to come after he has dealt with all serious objections and not merely with one of them.

Similarly, his reference to the principle of parsimony comes before its time. Use of the principle assumes, and is no substitute for, the

user's ability to show that entities have been multiplied beyond necessity. This ability can only be demonstrated by specific arguments directed, in the present case, at the claim that two different kinds of entity are needed. Having shown that *prime facie* one will do at least as well as two, we may be able to invoke the principle—but certainly not before. This requirement is not met by Deutscher's suggestion that "the physicalists' account of a mental state as that which has certain sorts of causes and which has certain sorts of effects is adequate to identify those brain states which exist when and only when the subject is in that mental state . . ."⁹ To find a psychophysical correlation is not the same as finding an identity of psychophysical properties. The former discovery alone no more puts it up to the dualist "to find something specifically nonphysical about the mental state"—as Deutscher says—than it puts to the physicalist the need to find something non-mental about the physical state. The burden of proof cannot be cast on to the antiphysicalist at *this* stage in the argument—with or without an appeal to the principle of parsimony.

But what of the various forms of the principle itself? Granted that the present example does not satisfy some of the conditions for appropriate use of the principle, can it ever properly be employed to decide between *philosophical* hypotheses and so to cast the burden of proof? Clearly, its use will be restricted to situations in which the competing philosophical hypotheses vary in their complexity. For if they do not, the principle will not enable us to prefer one to the other. If, then, we have a means of distinguishing between a simple and a less simple hypothesis, each of which explains all the presently available data, can we legitimately use the principle to cast *onus*? The answer, surely, is "Yes, but only if (a) we ever are in this position, and (b) we are able, as with any other principle, to justify its use if challenged, and in doing so to indicate the kinds of conditions in which it is applicable." Both conditions are more easily stated than exemplified.

Unless the principle in any of its versions and subforms—economy of entities, processes, and independent assumptions, simplicity of nature, simplicity of hypothesis—is interpreted merely as a maxim of terminological convenience, and thus irrelevant to our present purpose, the prin-

⁷ "Mental and Physical Properties" in C. F. Presley (ed.), *The Identity Theory of Mind* (Brisbane, 1967), pp. 74–75.

⁸ *Ibid.*, p. 70.

⁹ *Ibid.*, p. 75.

ciple implies that we should choose simplicity because of its substantive advantage. This advantage must be that the economical hypothesis is more likely to be true than the less simple one, given that both account for all the data so far produced. Our problem, then, is twofold: we must be able to recognize the simplest hypothesis; and we must be able to recognize which alternative hypotheses account for all the currently available data, that is, which hypotheses are relevant alternatives. Consider the second fold of the problem first.

No matter how it is stated, the principle of parsimony, in and of itself, provides us with no reason for choosing either between scientific hypotheses or between philosophical theories whose respective logical consequences are the same. Once we know the alternatives to be logically equivalent, there is no room left for either a methodological or ontological justification of parsimony. None of the alternatives is more likely to be true than any of the others. (Of course, if only one of the philosophical hypotheses entails the other their consequences will differ, and so the applicability of the principle will not come into question on that score.)

Similarly, the principle does not apply when the alternative hypotheses are logically independent in the sense that the truth or falsity of one does not entail the truth or falsity of the other. Given what we know in the test situation, they may all be true or all be false, or some be true and some false. Hence, choosing the simplest of logically independent hypotheses, even though all of them explain all the data so far produced, will not be a way of choosing the hypothesis most likely to be true. In this situation, parsimony will be no guide to truth, since we shall have no reason for believing only one, or even some, of the hypotheses most likely to be true while believing the remainder more likely to be false. If we wish to use the principle for selecting the hypothesis most likely to be true, we cannot do this when the choice must be made from logically independent alternatives. Thus we require that all the alternatives be inconsistent with each other for us to employ the principle. If any of the alternatives form a consistent set, the principle of parsimony can do no work for us in deciding between them.

What about the other part of the problem, that of our being able to recognize the simpler, or simplest, hypothesis? Obviously, the easiest case is that of two hypotheses which not only have a

stock of entities or processes in common, but differ in that one of the hypotheses uses additional entities or processes. Here we can sometimes argue that the additional factors are unnecessary because, as far as we now know, both hypotheses account for exactly the same data. But a difficulty seems to arise even here. For our opponent may reply that dispensing with superfluous entities or processes is not a *separate* advantage of the simpler hypothesis. Entities are superfluous if and only if their absence does not make the simpler hypothesis false: that is, if and only if their presence is not necessary for its truth. So we have to discover that the simpler hypothesis is true before we can determine that the extra entities are superfluous. Yet determining whether or not the simpler hypothesis is true in no way depends on discovering whether it employs fewer entities than its competitors. We ought to conclude, therefore, that the principle of parsimony, at least in this sort of case, is itself unnecessary. It records our achievement of truth rather than leading us to it.

Clearly, this answer, while correct in itself, is addressed to the wrong question. The principle of parsimony is supposed to indicate, *before* we know the truth, which sort of hypothesis is most likely to be true, namely, the one which uses fewer entities to explain the same data. The crucial problem, then, is how we can know at any given time that the competing hypotheses will be able to account for, well or badly, the same, increasing body of data in the future. Only if they do will they remain competitors. And only if they do, shall we be able, at some undetermined point, to identify the simplest hypothesis as the correct one, giving as our reason the sterility or the falsity of the various ancillary assumptions required to keep the other, and increasingly complex, hypotheses in the running. But all this lies in the future.

To show *now* that the additional entities are superfluous, we need to show now that the hypothesis of which they are a part can explain only the same body of data as the more frugal hypothesis. In science, this requires that we wait upon the results of mathematical treatment of the hypotheses, for it is only this treatment which will tell us whether the assumption of additional entities provides us with additional consequences that are testable in principle. Similarly in the case of competing philosophical hypotheses, we must wait upon the logical derivation of their respective consequences. Again, in both science and philo-

sophy we have to await the working out, by different means in the two fields, of the effects of these respective consequences upon other hypotheses and theories. It may be that such effects are few, but we cannot know this instantaneously. Establishing the logical relations between hypotheses takes time, precisely as the search for counter-examples to those hypotheses takes time. So while over a period of time we may be in a position to know whether a set of hypotheses can account for precisely the same body of data, it does not follow that at any given stage of the investigation we can, or will, know this. Hence, at any given stage we may or may not know that certain entities and processes are superfluous. That is, we shall not be able to identify the simpler, or simplest, hypothesis at some particular time by counting entities. And if we cannot do that, then we cannot use the principle of parsimony, on that occasion, to identify the hypothesis which is most likely to be true, namely, the simplest one.

But if this negative conclusion holds of the easiest case—that of hypotheses making use of a stock of common entities—what of the more difficult cases, those, for example, without entities in common, or those which make use of an equal number of independent assumptions? If counting entities is not sufficient to establish simplicity, how can we establish it when there is nothing to count? And, in particular, how can we establish it when the consequences of our hypotheses are largely unknown? We cannot, of course. The fact that the simpler hypothesis becomes increasingly probable, and the complex hypothesis increasingly improbable, with the accumulation of data explained by the former, need not be of practical help to us at any given time. It will not be unless we are familiar with the accumulated data. But if we are—if we have worked out the implications of the two hypotheses to discover whether they do explain the same collection of data—we no longer have any *predictive* use for the principle of parsimony. Even when the number of entities does turn out to be an accurate index of the relative simplicity of an hypothesis, we cannot know this early enough for the index to be of predictive value. By the time that our logical investigations have established that certain entities are superfluous, those same investigations have made the principle of parsimony itself superfluous.¹⁰ Hence, the principle can be used to ascribe *onus* only when

there is no genuine dispute over the relative simplicity of two hypotheses.

V

What do our examples show us? They show us, I suggest, that we were quite right in suspecting that the legitimacy of *onus*-assignment does not depend upon the acceptance of an identifiable set of philosophical propositions or principles which can prevail, *prima facie*, against all opposition. Short of principles of logic, there is, apparently, no such set—not even a one-member set. However, we were quite wrong in thinking that there might be some less general propositions with this power, and thus wrong in thinking that such propositions might ascribe *onus* at certain stages of particular arguments. For the generality of the candidates, as we have seen, has nothing to do with their supposed powers of *onus*-ascription. The reason is that there are no *onus*-assigning propositions of any sort. There are, it seems, only *onus*-assigning contexts or situations in which disputants find themselves, and in which they may legitimately lay a burden of proof upon one another.

What is accepted or rejected in a given debate is the specific use to which a certain proposition or principle is put. The plausibility of that specific use depends upon the other premisses and conclusions with which the proposition or principle in question is conjoined. Since in theory any of them is open to attack, it is often no defense of a proposition to claim that it casts a burden of proof on its attacker. The premisses by which this ascription is supported can themselves be rejected. But in advance of a given discussion we cannot know, except in the most general and unhelpful way, which propositions and principles all our possible opponents will retain or reject. Hence, we cannot usefully put forward propositions and principles which will lay a burden of proof against all future disputants. Yet that is what we attempt to do when we advance the claims of particular propositions in ignorance of our opponents' countering moves. If we describe a specific set of such moves—a set of propositions or principles forming one side of a debate—then given this set we can legitimately ascribe *onus*. It was the absence of such a set from each of our examples that allowed us to point out the difficulty of attaching *onus* to propositions when their environment is unknown.

¹⁰ The argument of the last two paragraphs is a brief adaptation of that by George Schlesinger in *Method in the Physical Sciences* (London, 1963), pp. 32–39.

The same difficulty can arise when the philosophy lecturer, after going through Moore's proof for the existence of an external world, says "Now the burden is on you to show that this is not my right hand and that it is not a material thing." For sometimes the students reply "No, the *onus* is on you to prove both points. Why should it lie on us?" The answer, obviously, is that the *onus* now lies on them to show why this object held up is not the lecturer's hand. They have been given evidence. Is it not good enough for them? Do they maintain that they are victims of optical trickery, mass hallucination, or auditory error? And that, therefore, what is being held up is not the lecturer's hand, or not his right hand, or that there is no lecturer in the room? Or do the students maintain that he is holding up his right hand but that his particular hand is not a material object because no hands are material objects? When the lecturer learns which claims the students take to be in doubt he will also learn what is to be proved. When he knows that, he will be in a position to judge the plausibility of the supporting argument. Yet the initial burden of proof still lies on the side with the weaker evidence—in the present case, that of the students.

But if this is all that the assignment of *onus* comes to, then it has very limited uses as a weapon in argument. When most wanted—to overthrow an opponent who rejects an apparently conclusive argument—it cannot be employed, since *onus* cannot be forced on a clear-headed but unwilling antagonist. For he can always appeal in defense to a

difference in premisses, if his other views permit this. And if they do not, then laying the burden of proof on him is merely to point out what his own arguments commit him to establishing, and sometimes to indicate his errors in reasoning and his trifling with evidence. So that if a philosopher is clear-headed he lays the burden of proof on himself. In a successful discussion each participant recognizes this, and his arguments are designed to discharge his burden. If the *onus* is laid on him by others, and he agrees, it will be because he recognizes the logical powers of the propositions, and the weight of evidence of the claims, to which he is committed. *Onus*-ascription, then, is backward-looking, not forward-looking. It is the register and summary of agreement achieved, or assumed, rather than being a reliable prediction of the future course of a dispute. Explicitly raising the question of *onus* is simply a way of flagging the progress of an argument.

Hence, the problem of *onus*-assignment is part and parcel of the questions, "What counts as an acceptable argument in this field? What sorts of premisses are plausible? What sorts of considerations are telling? What sorts of procedural principles are allowable?" When the participants share enough answers to these questions so that they are able to discuss a given topic, and not be thrown back on to more fundamental disagreement, then they can also agree on assignment of the burden of proof; otherwise its ascription is not defensible, and certainly not labor-saving.¹¹

Australian National University and University of Massachusetts

Received February 28, 1969

¹¹ I am grateful to Professors John Passmore and Judith Jarvis Thomson for their criticism.

VII. A NON-CLASSICAL THEORY OF TRUTH, WITH AN APPLICATION TO INTUITIONISM

STORRS McCALL

IT is characteristic of what I shall call the "classical" theory of truth that it satisfies Tarski's criterion, to the effect that ' p ' is true if and only if p , and that it satisfies the principle of bivalence, to the effect that every proposition is either true or false. In a previous paper, I have made note of some of the consequences of rejecting the principle of bivalence in the field of propositions concerning the future.¹ What I now propose to do is to set down some of the formal properties of a nonclassical theory of truth in greater detail, and to indicate another and quite different field of application.

To begin with, it is worth noting that rejection of the principle of bivalence entails rejection of Tarski's criterion. Strictly speaking, it should really be to Aristotle rather than to Tarski that this criterion is attributed, in view of the famous statement in *Metaphysics* 1110b26-28 that "to say of what is that it is not, or of what is not that it is, is false; while to say of what is that it is, or of what is not that it is not, is true." Furthermore, it is to Aristotle that we attribute the heroic rejection of the principle of bivalence in connection with propositions concerning future events.² What Aristotle seems not to have noticed is that the two doctrines, that ' p ' is true if and only if p , and that some propositions are neither true nor false, are incompatible.

To see this, let a be a proposition that is neither true nor false:

$$NTa \ \& \ NFa^3$$

Since there is no difference in meaning between saying that p is false and saying that Np is true⁴, we have $Fp \equiv TNp$, which yields:

$$NTa \ \& \ NTN\alpha$$

Applying Tarski's equivalence, we obtain

$$Na \ \& \ NN\alpha,$$

which is a contradiction. Hence either Tarski's criterion or the doctrine that some propositions are neither true nor false must be discarded.⁵ The orthodox solution, of course, is to retain Tarski's criterion together with the principle of bivalence—these two constitute central pillars of the classical theory of truth. But it is not difficult to see that a very different theory would result from their rejection.

In setting forth in the next few paragraphs the outline of a nonclassical theory of truth, our main purpose will be to see what formal properties are possessed by the expressions "it is true that . . ." and "it is false that . . .," represented by the operators T and F . In so doing we shall proceed in a quite intuitive way, separating out those propositions which seem to state truths about truth and falsehood from those which do not. The result

¹ "Temporal Flux," *American Philosophical Quarterly*, vol. 3 (1966), pp. 270-281.

² According, that is, to one interpretation of what Aristotle is saying in *De Interpretatione* ix. Another interpretation (which I consider less well supported) is that all propositions are either true or false, but that in addition all propositions about the past and present are necessarily (=unpreventably) true, whereas not all propositions about the future are. For these two possible interpretations see J. L. Ackrill, *Aristotle's "Categories" and "De Interpretatione"* (Oxford, 1963), pp. 139-142; and Nicholas Rescher, "Truth and Necessity in Temporal Perspective" in Richard Gale (ed.), *The Philosophy of Time* (New York, 1967), pp. 184-194.

³ We use T for "it is true that," F for "it is false that," N for "it is not the case that," and Peano-Russell or Hilbert notation for the rest.

⁴ Intuitionists would disagree with this. See the remarks at the end of the paper on intuitionist negation.

⁵ Passages in which the rejection of Tarski's criterion is mooted may be found in Michael Dummett's "Truth," *Proceedings of the Aristotelian Society*, vol. 59 (1958-59), pp. 141-162, and in Nicholas Rescher's "On the Logic of Chronological Propositions," *Mind*, vol. 75 (1966), p. 78.

will be a formal system—a surprisingly familiar one at that—with axioms and rules of inference embodying the sought-for nonclassical properties of truth and falsehood.⁶ In fact two such formal systems suggest themselves, as will be seen.

In constructing a nonclassical theory of truth, one of the first things to note is that rejection of the principle of bivalence does not entail rejection of the law of the excluded middle. These two are in fact very different, although they have often been confused. In the symbolism introduced above, the principle of bivalence is formulated as follows:

$$(1) Tp \vee \neg Tp$$

while the law of the excluded middle is:

$$(2) p \vee \neg p.$$

The first philosopher in modern times explicitly to make the distinction between the two was Jan Łukasiewicz, although the difference has been subsequently blurred by the fact that Łukasiewicz's own three-valued logic, specifically constructed so as to violate the principle of bivalence and allow for propositions which are neither true nor false, also violates the law of the excluded middle.⁷ Of course if Tarski's criterion were accepted, the difference between the two would vanish. Applying the equivalence $Fp \equiv \neg Tp$ to (1) yields

$$(3) Tp \vee \neg Tp,$$

which reduces to (2) if we hold that $Tp \equiv p$. But in the absence of the latter we are free to reject the principle of bivalence without thereby rejecting the law of the excluded middle: we shall in fact retain this law as one of the elements in our nonclassical theory of truth.

Care is called for, however. It is not implausible to assume that if the law of the excluded middle holds, it is true:

$$(4) \vdash T(p \vee \neg p),$$

⁶ Note that there will be no danger of inconsistency, through formulation of the liar paradox, in this insertion of an analogue for "true" in a formal system, since the system in question is not semantically closed. See Alfred Tarski, "The Semantic Conception of Truth," *Philosophy and Phenomenological Research*, vol. 4 (1944), pp. 341-375, reprinted in L. Linsky (ed.), *Semantics and the Philosophy of Language* (Urbana, 1952). The approach adopted in this paper is quite different from Bas van Fraassen's in "Presupposition, Implication and Self-reference," *The Journal of Philosophy*, vol. 65 (1968), pp. 136-152, in which a nonclassical theory of truth is constructed for the purpose of resolving the semantic paradoxes. Nevertheless, the resulting nonclassical theories are quite similar.

⁷ Łukasiewicz makes the distinction in "Philosophische Bemerkungen zu mehrwertigen Systemen des Aussagenkalküls," *Comptes rendus des séances de la Société des Sciences et des Lettres de Varsovie*, Class III, vol. 23 (1930), pp. 51-57. Translated in Storrs McCall (ed.), *Polish Logic: 1920-1939* (Oxford, 1967). In "Temporal Flux," *op. cit.*, p. 277, I show how a small change in the definition of the notion of disjunction in Łukasiewicz's three-valued logic would enable him to assert the law of the excluded middle while at the same time deny the principle of bivalence. Hence the two are logically independent.

⁸ W. V. Quine, "On a So-called Paradox," *Mind*, vol. 62 (1953), p. 65.

and if we were also to hold that $T(p \vee q)$ implied $Tp \vee Tq$, we would be able to derive (3) from (4). Hence we must reject the thesis corresponding to this implication:

$$(5) \vdash T(p \vee q) \supset (Tp \vee Tq),$$

thus embracing what Quine calls "Aristotle's fantasy that 'It is true that p or q ' is an insufficient condition for 'It is true that p or it is true that q .'"⁸ Aristotle's fantasy though, so far from being the product of an overheated imagination, emerges as no more than a sober means of blocking the derivation of the principle of bivalence from the law of the excluded middle.

Turning now to Tarski's criterion, the equivalence $Tp \equiv p$ is separable into the two implications $Tp \supset p$ and $p \supset Tp$. The second of these would again allow passage from (2) to (3) via the law $[(p \vee q) \& (p \supset r) \& (q \supset s)] \supset (r \vee s)$, and so must be rejected:

$$(6) \vdash p \supset Tp.$$

The first, however, entails no such unwelcome consequences and in fact seems to make a correct assertion about truth: if it is true that snow is white, then snow is white. Thus we have:

$$(7) \vdash Tp \supset p.$$

Note that the effect of (6) and (7) is to make truth something stronger than bare assertion; to say "it is true that snow is white" is to say something stronger than merely "snow is white," in the sense that the second can be inferred from the first but not vice versa. This difference in strength seems to be a necessary feature of the nonclassical theory of truth.

Although it is impossible to accept without contradiction both limbs $Tp \supset p$ and $p \supset Tp$ of the equivalence $Tp \equiv p$, and the doctrine that some propositions are neither true nor false, nevertheless a weaker version of (6) is perfectly harmless. This is the rule of inference entitling us to pass

from any true proposition a to the proposition "it is true that a ":

$$(8) \vdash a \rightarrow \vdash Ta$$

Rule 8 permits us to argue, for example: "Snow is white, therefore it is true that snow is white," while by contrast (6) forbids us to say: "If snow is blue then it is true that snow is blue." We accept the rule of inference, while rejecting the corresponding implicative thesis.

A different implicative thesis about truth, on the other hand, is the following:

$$(9) \vdash T(p \supset q) \supset (Tp \supset Tq).$$

(9) states in effect that if both 'if p then q ' and ' p ' are true, then ' q ' is true; an analogue of the rule of *modus ponens* whose correctness it seems impossible to deny.

Finally, a group of important formal properties of truth and falsehood concerns iteration of the operators T and F . To begin with, iteration of the truth-operator yields nothing new, since to say "it is true that it is true that p " means simply "it is true that p ":

$$(10) \vdash TTp \equiv Tp.$$

Similarly, "it is true that it is false that p " means only "it is false that p ";

$$(11) \vdash TFp \equiv Fp.$$

But the situation is quite different when it is the falsehood-operator that is iterated. "It is false that it is false that p " means neither "it is false that p " nor "it is true that p " (in the latter case p might be neither true nor false, so that its not being false would not entail its being true). Instead, to say "it is false that it is false that p " is to say something weaker than p . We have:

$$(12) \vdash FFp \equiv Fp$$

$$(13) \vdash p \supset FFp$$

$$(14) \vdash FFp \supset p.$$

Comparison of (12)–(14) with (10)–(11) reveals the enormous structural difference between truth and falsehood.

The fourth possible combination of truth-operators, FTp , raises difficulties. Plainly "it is false that it is true that p " is entailed by "it is false that p ," and (still plausibly) by "it is not the case that p ":

$$(15) \vdash Fp \supset FTp$$

$$(16) \vdash Np \supset FTp.$$

But is FTp implied by "it is not true that p "? The issue is a fairly important one, since what is at stake is the question of which of two possible well-known formal systems fits our nonclassical theory of truth best.

For modal logicians there will be something very familiar about the truth-operator T and the properties that have up to now been given it. In all formal respects it is precisely analogous to the necessity-operator L (or \Box) in Lewis' system S_4 . Defining F as TN , and replacing T by L throughout, we find that formulae (2), (4), (7–11), (13), (15), and (16) all hold in S_4 , while (1), (3), (5), (6), (12), and (14) fail. Among the first group are to be found all the axioms and rules of Gödel's axiomatization of S_4 , it being understood that the underlying non-modal propositional logic is two-valued.⁹ But now for the first time, in considering whether "it is not true that p " implies "it is false that it is true that p " we are faced with a proposition which holds in S_5 but not in S_4 . As the matter of whether the proposition should be accepted seems not to be clear-cut, the best thing is to leave the question open:

$$(17) NTp \supset FTp.$$

If (17) is accepted, the formal system corresponding to our nonclassical theory of truth will be S_5 , since (17) is a version of S_5 's characteristic axiom. If not it will be S_4 . In either case we have a system embodying a concept of truth for which Tarski's criterion and the principle of bivalence fail.

Before proceeding to a possible application of the theory sketched here, comparison should be made between it and Łukasiewicz' three-valued logic, also designed to accommodate propositions neither true nor false. Łukasiewicz assigned such propositions a *third truth-value*, to which he originally gave the description "indifferent" or "indeterminate," and which he eventually came to characterize as "possible." Furthermore, his logic is *truth-functional* in the sense that the truth-value of any complex proposition constructed with the help of propositional connectives is uniquely determined by the truth-values of its component propositions. The theory of this paper differs from Łukasiewicz' in that it is noncommittal as to the existence of a third truth-value. That is, it is quite consistent with the theory that there should be only *two* truth-values, namely "true" and "false," and that there should be at the same time propositions which take

⁹ See for example A. N. Prior, *Formal Logic* (Oxford, 1955), p. 306 for this axiomatization.

neither of these values. The point at issue is not *quite* a verbal quibble: after all, why should "neither true nor false" be thought to be a truth-value any more than "either true or false," or for that matter "both true and false?" Secondly, and more importantly, the system proposed here differs from Łukasiewicz' in not being truth-functional. And this (in its author's eyes) is one of its main virtues: it has always seemed a purely arbitrary matter why in Łukasiewicz' logic a conditional proposition with a "true" antecedent and an "indeterminate" consequent should itself receive the value "indeterminate" rather than some other value. Or (perhaps better) rather than a value which is totally independent of the truth-values of its components, as might well be the case if its antecedent and consequent were irrelevant to one another in content. Therefore, quite apart from the matter of whether we wish to call "neither true nor false" a truth-value, the most important respect in which the new theory differs from Łukasiewicz' is in not being truth-functional. For convenience, let us call it a non-truth-functional three-valued logic.

Now for an application. One of the areas that comes to mind in which a theory of truth that rejects the principle of bivalence might be of use is the realm of the undecidable in mathematics. For coping with the undecidable (or more commonly, as their examples show, with a multitude of problems no decision procedure for which is known, and which therefore *may* be undecidable) the intuitionists have devised a very ingenious logic of propositions. Central to this logic, as is well known, is the rejection of the law of the excluded middle and a doctrine of negation which results in the failure of the law of double negation elimination $\neg\neg p \supset p$. What I shall try to show is that an alternative way of dealing with the kinds of problem discussed by intuitionists is to reject the principle of bivalence. This solution has the virtue of making it no longer necessary to reject excluded middle and double negation, and a more orthodox propositional logic may be used instead.

The best way of seeing how this works out in practice is to consider one or two of the examples frequently employed by intuitionists. These examples concern methods of reasoning in mathematics which present no problem when applied to finite collections, but which yield what the

intuitionists claim are unjustified and illegitimate results when applied to infinite collections. Consider for example a random sequence of digits which though very long is finite. It is perfectly clear how we should go about determining whether, say, an uninterrupted run of 100 sevens occurs in this sequence: we merely go through (or employ a computer to go through) the sequence one by one. But when the sequence is infinite the situation is different. If we ask whether the decimal development of π contains an uninterrupted run of 100 sevens, it is not at all clear how to search for an answer. Enumerating the decimals one by one might conceivably yield an affirmative answer (though in that case one could ask exactly the same question about a run of 1,000 sevens), but would never yield a negative answer. But must the question have either an affirmative or a negative answer? That it must is indicated by the following argument, which though classically valid is rejected by the intuitionists.

Write down the decimal development of π , and write underneath it the decimal d :

$$\pi = 3.141592653 \dots$$

$$d = 0.33333333 \dots$$

where d consists entirely of threes up to the end of the first occurrence of an uninterrupted run of 100 sevens in the decimal development of π , and of zeros thereafter. Plainly the question as to whether or not 100 sevens occur in π has an answer if and only if there exists a unique real number D corresponding to the decimal d . (Thus if 100 sevens occur at a certain specific place in π , D is a specific real number less than $1/3$ and conversely; if 100 sevens never occur $D = 1/3$ and conversely.¹⁰) That there is such a unique real number D is provable as follows:

- (i) Assume that there is no number n which is the decimal place of the termination of the first uninterrupted run of 100 sevens in the decimal development of π . Then $D = 1/3$.
- (ii) Assume that there is such a number n .

$$\text{Then } D = 0.3333 \dots 3 = \frac{10^n - 1}{3 \cdot 10^n}$$

- (iii) But either there is such a number n or there is not. There being no possible third alternative, the number D exists and is unique.

¹⁰ Is D a rational number? One would think so, but the intuitionists' view of the matter is that although one can say that it is not the case that D is not rational, one cannot say that D is rational. See Arend Heyting, *Intuitionism* (Amsterdam, 1966, 2nd ed.), p. 17 for a discussion of this.

Intuitionists dissociate themselves from this argument because of step (iii), which is an instance of the law of the excluded middle. But it is precisely on the law of the excluded middle that a classical mathematician would insist most strongly, maintaining that nothing would ever shake him in his conviction that either there is a number n which is the decimal place of the termination of the first uninterrupted run of 100 sevens in the decimal development of π , or there is not. What I am suggesting is that the intuitionists *give* him the law of the excluded middle as an undisputed logical truth, and instead claim that the statement about 100 sevens in π is not subject to the principle of bivalence. It seems quite plausible to maintain that if it is in principle undecidable whether or not there are 100 sevens in π (as it may well be), then it is neither true nor false that there are. And, as will be seen, nothing need be lost in the argument.

The chain of reasoning concerning the decimal D above took the following form:

- (i) Assume p . It follows that q (i.e. that D is unique).
- (ii) Assume $\neg p$. It follows that q .
- (iii) But p or $\neg p$.
- (iv) Hence q .

What exactly is being assumed in line (i)? That p , yes, but what if p is a proposition that is neither true nor false? It seems hard to avoid saying that what is being assumed in line (i) is that p is true. Similarly, in line (ii), that p is false. But then, of course, a possibility still remains, namely that p is neither true nor false, and cases (i) and (ii) are not exhaustive. For this reason the conclusion q is not derivable, even with the help of the law of the excluded middle (iii). A valid version of the constructive dilemma would have to look something like this:

- (i) Assume Tp . It follows that q .
- (ii) Assume Fp . It follows that q .
- (iii) Assume NTp & NFp . It follows that q .
- (iv) But Tp or Fp or $(NTp \& NFp)$.
- (v) Hence q .

Here the general rule to be applied to arguments from hypotheses would be: Never make an assumption without making it clear whether what is assumed is to be taken as true, false, or neither.

Adherence to this rule, and rejection of the principle of bivalence as applied to problems for which no decision method is known, provides an alternative and, I think, more acceptable way of avoiding conclusions unwelcome to intuitionists.

One other type of proof which is intuitionistically suspect is *reductio ad absurdum*, the schema of which is as follows:

- (i) Assume $\neg p$.
- (ii) It follows that, for some q , $q \& \neg q$.
- (iii) Hence p .

Here again, if p is an undecidable proposition the fact that $\neg p$ implies a contradiction, and hence that p cannot be false, does not in itself imply that p must be true, since p may be neither true nor false. Consider the following counter-example, due to E. W. Beth:¹¹

If there is a natural number with the property A , then there is a smallest natural number with the property A .

Proof (using *reductio ad absurdum*). Let n have the property A , and assume there is no smallest number which possesses A . Then there must be a number $n_1 < n$ possessing A , since otherwise n would be the smallest number possessing A . Likewise there must be an $n_2 < n_1$ possessing A , in fact there must be a descending progression of numbers

$$n > n_1 > n_2 > \dots$$

all with property A . This progression contains at most n terms and thus must terminate with a certain number m . But it follows that m is the smallest number with the property A , contrary to the hypothesis that there is no such number. This is a contradiction. Hence there is a smallest number with property A .

"So far so good," the reader may say. "I see nothing wrong with this." But wait. According to the intuitionists the theorem proved above is untrue, as may be seen by taking for A the following property:

The natural number n has the property A iff there are at least $100-n$ perfect numbers.¹²

The numbers 99, 98, and 97 have the property A , and hence by the theorem there must be a smallest number having the property A . But at the same time only 12 perfect numbers are known to

¹¹ E. W. Beth, *Mathematical Thought* (Dordrecht, 1965), p. 82. Beth's *Foundations of Mathematics* (Amsterdam, 1959), also contains interesting intuitionistic examples.

¹² A perfect number is equal to the sum of all its proper factors: e.g., $6 = 3 + 2 + 1$; $28 = 14 + 7 + 4 + 2 + 1$.

exist, and the question as to how many there are altogether may be undecidable. In that case it would be impossible to determine the smallest number n having the property A . So what do we make of the proof that there is such a number? Well, if we wish we can take the hard line of saying that while there is no formal contradiction here (since we do not have a proof that there is no smallest number n) nevertheless if it is impossible to find such an n , to say that we have proved that it exists is absurd.¹³ Or we can take the softer line of saying that the statement that such an n exists "has very little intuitive content."¹⁴ In either case serious doubt seems to be cast on the validity of proofs by *reductio ad absurdum*.

I have already indicated how *reductio* proofs rely implicitly on the principle of bivalence. I now want to call attention to a peculiar but well-known asymmetry in intuitionist logic. Whereas the *reductio* argument stated earlier in which the assumption Np occurred was invalid, the following argument is valid:

- (i) Assume p
- (ii) It follows that, for some q , $q \& Nq$.
- (iii) Hence Np .

In other words, from the falsehood of p we can infer Np , but from the falsehood of Np we cannot infer p . Why this difference? The answer lies in the idiosyncracies of the intuitionist doctrine of negation. In classical logic negation transforms a proposition p into a proposition Np with the property that if one of the two is true the other is false, and vice versa. This is not so in intuitionist logic, where although p and Np cannot both be true, they can both be false. That is, p and Np behave more like *contraries* than like *contradictories*, and in fact intuitionists frequently read " Np " as " p is absurd" rather than " p is false." More precisely, in intuitionist logic the following principle holds:

- (a) If p is true (false), then Np is false (true), as is reflected in the law of double negation introduction $p \supset NNp$ (if p is true, then Np is false; if Np is false, NNp is true; hence if p is true, NNp is

true). But the following principle of classical negation does *not* hold:

- (b) If Np is true (false), then p is false (true), since if it did the law of double negation elimination $NNp \supset p$ would hold. It is probable that these irregularities in the doctrine of negation have caused more anguish to those trying to understand intuitionism than any other feature of intuitionist logic.

By substituting rejection of the principle of bivalence for rejection of the law of the excluded middle, the irregularities in intuitionist negation can be entirely done away with. Furthermore, the asymmetry involved in accepting the species of *reductio* argument which assumes the hypothesis p , while rejecting that which assumes the hypothesis Np , may also be discarded. Both in fact will be equally invalid (since they make implicit appeal to the principle of bivalence), although the following will take their place as a valid form of *reductio* argument:

- (i) Assume Fp . It follows that, for some q , $q \& Nq$.
- (ii) Assume $NTp \& NFp$. It follows that, for some q , $q \& Nq$.
- (iii) Hence Tp .

Goodstein (*op. cit.*, p. 5) tells the following story about Brouwer, who began to criticize *reductio ad absurdum* proofs in 1913. Brouwer, who was at that time one of the editors of *Mathematische Annalen*, "opened the attack by rejecting all papers offered to the *Annalen* which applied the *tertium non datur* to propositions the truth or falsehood of which could not be decided in a finite number of steps. The Editorial Board met this emergency by resigning—and then re-electing themselves, minus Brouwer. Incidentally, the Dutch Government so resented this slight on their leading mathematician that they founded a rival mathematical journal, with Brouwer in charge." One wonders whether the editors of *Annalen* would have been mollified had Brouwer contented himself with rejecting the universal applicability of the principle of bivalence, and left the *tertium non datur* alone.

University of Pittsburgh

and

Makerere University College, Kampala, Uganda

Received October 30, 1968

¹³ See R. L. Goodstein, *Essays in the Philosophy of Mathematics* (Leicester, 1965), p. 4.

¹⁴ Beth, *op. cit.*, p. 83.

BOOKS RECEIVED

- BRANDT, Richard B. (ed.), *Social Justice*. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1962. Pp. vi+171. (This anthology contains excerpts from the following authors: Kenneth E. Boulding, Paul A. Freund, William K. Frankena, Alan Gewirth and Gregory Vlastos.) Paper, \$1.95.
- FODOR, Jerry A., *Psychological Explanation: An Introduction to the Philosophy of Psychology*. New York: Random House, 1968. Pp. ix+165. Paper, \$1.95.
- GIBSON, A. Boyce, *Muse & Thinker*. London: C. A. Watts & Co., 1969. Pp. ix+177. (Published in The New Thinker's Library series.) \$1.80.
- GOULD, James A. and Vincent V. Thursby (eds.), *Contemporary Political Thought: Issues in Scope, Value, and Direction*. New York: Holt, Rinehart and Winston, 1969. Pp. viii+376. (This anthology contains excerpts from the following authors: George H. Sabine, George E. G. Catlin, Leo Strauss, William A. Glaser, Andrew Hacker, Robert A. Dahl, Christian Bay, T. D. Weldon, Joseph Margolis, W. G. Runciman, David Easton, Paul F. Kress, Jean-Paul Sartre, T. L. Thorson, Gabriel A. Almond, Alfred Cobban, Isaiah Berlin, and Harry Eckstein.) Paper, \$4.95.
- GUITTON, Jean, *La pensée et la guerre*. Bruxelles: Desclée de Brouwer, 1969. Pp. 228.
- ISEMINGER, Gary, *An Introduction to Deductive Logic*. New York: Appleton-Century-Crofts, 1968. Pp. vii+184. Published in The Century Philosophy Series. Paper, \$2.95.
- (ed.), *Logic and Philosophy*. New York: Appleton-Century-Crofts, 1968. Pp. viii+248. (This anthology contains excerpts from the following authors: Gottlob Frege, A. D. Woozley, George Pitcher, Bertrand Russell, Hans Hahn, Arthur Pap, W. V. Quine, C. I. Lewis, Alan Ross Anderson and Nuel D. Belnap Jr., Alexius Meinong, Morton White, William P. Alston, Czeslaw Lejewski, L. Jonathan Cohen, P. F. Strawson, J. A. Farris, and Franz Brentano.) Published in The Century Philosophy Series. Paper, \$3.50.
- JOHNSTON, Frederick S., Jr., *The Logic of Relationship*. New York: Philosophical Library, 1968. Pp. 110. \$4.00.
- KAUFMANN, Walter, *Nietzsche: Philosopher, Psychologist, Antichrist*. New York: Vintage Books, 1968. Pp. xviii+524. (This is a third edition, revised and enlarged.) Paper, \$2.45.
- KLEMKE, E. D., *The Epistemology of G. E. Moore*. Evanston, Ill.: Northwestern University Press, 1969. Pp. xi+205.
- KLIBANSKY, Raymond (ed.), *Logic and Foundations of Mathematics*. Firenze: La Nuova Italia Editrice, 1968. Pp. xi+387. (This anthology contains excerpts from the following authors: Jaakko Hintikka, Nicholas Rescher, Henry Veatch, R. M. Martin, Abraham Robinson, Ruth Barcan Marcus, Richard Montague, Paul Lorenzen, G. H. von Wright, Ch. Perelman, Henry W. Johnstone Jr., Tadeusz Kubiński, Jerzy Stupecki, Maria Kokoszyńska, A. A. Zinovjev, G. C. Moisil, Ettore Casari, Shoji Maehara, Alonzo Church, G. Hasenjaeger, Kurt Gödel, Hans Hermes, Akiko Kino, Hilary Putnam, Saunders MacLane, Haskell B. Curry, Arend Heyting, John Myhill, and H. Freudenthal.) Published in the Contemporary Philosophy Series.
- KRAUS, John Louis, *John Locke: Empiricist, Atomist, Conceptualist and Agnostic*. New York: Philosophical Library, 1968. Pp. 202. \$4.95.
- KYBURG, Henry E., Jr., *Probability Theory*. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1969. Pp. x+294. \$10.95.
- LYMAN, Frederick C., Jr., *The Posture of Contemplation*. New York: Philosophical Library, 1969. Pp. 123. \$3.95.
- MARGOLIS, Joseph (ed.), *An Introduction to Philosophical Inquiry*. New York: Alfred A. Knopf, 1968. Pp. xii+942. (This anthology contains excerpts from the following authors: C. D. Broad, Rudolf Carnap, John Wisdom, St. Anselm, St. Thomas Aquinas, Immanuel Kant, Søren Kierkegaard, A. J. Ayer, Paul Tillich, C. B. Martin, J. L. Mackie, St. Augustine, John Stuart Mill, C. A. Campbell, Stuart Hampshire, R. S. Peters, Donald Davidson, H. L. A. Hart and A. M. Honoré, Aristotle, David Hume, Ernest Nagel, Carl G. Hempel, René Descartes, George Berkeley, Gilbert Ryle, J. J. C. Smart, P. T. Geach, Hilary Putnam, P. F. Strawson, Bertrand Russell, Norman Malcolm, G. E. M. Anscombe, Hans Reichenbach, Nelson Goodman, W. V. Quine, William James, J. L. Austin, C. S. Peirce, Moritz Schlick, Friedrich Wais-

- mann, Gottlob Frege, D. F. Pears, Plato, John Locke, G. E. Moore, O. K. Bouwsma, Roderick M. Chisholm, Thomas Hobbes, F. H. Bradley, John Dewey, Brian Medlin, R. B. Perry, and Kurt Baier.) \$8.95.
- MOLINA, Fernando R. (ed.), *The Sources of Existentialism as Philosophy*. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1969. Pp. x+230. (This anthology contains excerpts from the following authors: Søren Kierkegaard, Friedrich Nietzsche, Edmund Husserl, Martin Heidegger, José Ortega y Gasset, Jean-Paul Sartre, Paul Tillich, Eugene T. Gendlin, and Maurice Merleau-Ponty.) Paper, \$2.95.
- RADCLIFF, Peter (ed.), *Limits of Liberty: Studies of Mill's On Liberty*. Belmont, Calif.: Wadsworth Publishing Co., 1966. Pp. viii+118. (This anthology contains excerpts from the following authors: Albert William Levi, Alexander Meiklejohn, Willmoore Kendall, James Fitzjames Stephen, H. L. A. Hart, J. S. Mill, Isaiah Berlin, S. I. Benn and R. S. Peters, J. C. Rees, and Marcus George Singer.) Published in the Wadsworth Studies in Philosophical Criticism. Paper, \$4.67.
- RAO, P. Nagaraja, *Values in the Changing World*. Bangalore: The Indian Institute of World Culture, 1968. Pp. 41. (Transaction No. 37.) Paper.
- RESCHER, Nicholas, *Introduction to Value Theory*. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1969. Pp. viii+199. Paper, \$2.95.
- , *Topics in Philosophical Logic*. New York: Humanities Press, 1969. Pp. xiv+347. \$18.50.
- ROBISON, Gerson B., *An Introduction to Mathematical Logic*. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1969. Pp. xi+212. \$5.95.
- ROSS, James F., *Philosophical Theology*. New York: Bobbs-Merrill, 1969. Pp. x+326. \$8.50.
- ROTENSTREICH, Nathan (ed.), *The Understanding of History*. Jerusalem: The Israel Academy of Sciences and Humanities, 1968. Pp. 214. (This book contains excerpts from the following authors: K. Löwith, N. A. Nikam, R. P. McKeon, Ch. Perelman, A. G. M. van Melsen, J. Hyppolite, and E. Topitsch.)
- SCHMITT, Richard, *Martin Heidegger on Being Human*. New York: Random House, 1969. Pp. 274. Paper, \$2.95.
- TILLMAN, Frank A. and Steven M. Cahn, *Philosophy of Art and Aesthetics: From Plato to Wittgenstein*. New York: Harper & Row, 1969. Pp. xi+791. \$14.00.
- VEATCH, Henry B., *Two Logics*. Evanston, Ill.: Northwestern University Press, 1969. Pp. viii+280. \$8.00.
- WESTLING, Achilles, *On the Constitution of Experimental Knowledge*. Helsinki: distributed by The Academic Book Store, 1968. Pp. 59. Paper.
- WIPPEL, John F. and Allan B. Wolter, O.F.M. (eds.), *Medieval Philosophy: From St. Augustine to Nicholas of Cusa*. New York: The Free Press, 1969. Pp. viii+487. Published in Readings in the History of Philosophy Series. Paper, \$3.95.
- WISDOM, John, *Other Minds*. Berkeley and Los Angeles: University of California Press, 1968. Pp. viii+265. (Originally published by Basil Blackwell in 1952 and 1965.) Paper, \$2.25.

American Philosophical Quarterly Monograph Series

Edited by NICHOLAS RESCHER

Studies in Moral Philosophy

1. On Moral Truth KAI NIELSEN
2. On Ethical Egoism JESSE KALIN
3. Moral Nihilism G. P. HENDERSON
4. Supererogation and Duties MICHAEL STOCKER
5. Utility and Rights LAWRENCE HOWARTH
6. Let Needs Diminish that Preferences may Prosper DAVID BRAYBROOKE
7. Whewell's Ethics JEROME B. SCHNEEWIND

Med. 8vo, 144 pp., paper 35s. net 63s 11450 5

Studies in Logical Theory

1. Two Types of Denotation MONTGOMERY FURTH
2. Language Games for Quantifiers JAAKO HINTIKKA
3. Types, Categories and Non-sense JAMES W. CORNMAN
4. A Theory of Conditionals ROBERT C. STALNAKER
5. Goodman's Nominalism ALAN HAUSMAN and CHARLES ECHELBERGER
6. Truth: Austin, Strawson, Warnock TED HONDERICH
7. Propositions and Abstract Propositions COLWYN WILLIAMSON

Med. 8vo, 152 pp., paper 35s. net 63s 11460 2



BASIL BLACKWELL, Oxford, England

DIALOGUE

Canadian Philosophical Review—Revue Canadienne de Philosophie

Editors: VENANT CAUCHY and MARTYN ESTALL

VOL. VIII—1969—No. 3

ARTICLES

- | | |
|--|------------------|
| Senses of Identity in Hume's <i>Treatise</i> | JAMES NOXON |
| Le Problème de la perception chez Leibniz | YVON BÉLAVAL |
| La Révolution copernicienne: Freud et le géocentrisme médiéval | CLAUDE SAVARY |
| On a Bill of Rights | R. N. McLAUGHLIN |
| Le Rapport âme-corps chez le premier Marcel | GERMAINE CROMP |
| Weinberg's Refutation of Nominalism | FRED WILSON |

NOTES

DISCUSSIONS

REVIEWS

Subscriptions: \$7.50 a year to individuals: \$8.00 to Libraries.
Payable to the Canadian Philosophical Association in care of H. M. Estall: Department of Philosophy, Queen's University, Kingston, Ontario, Canada.

THE REVIEW OF METAPHYSICS

A PHILOSOPHICAL QUARTERLY
Edited by RICHARD J. BERNSTEIN

Volume XXIII, No. 2 Issue No. 90

December, 1969

ARTICLES:

- | | |
|------------------|--|
| Walter Kaufman | <i>The Origin of Justice</i> |
| Geoffrey Hartman | <i>The Voice of the Shuttle</i> |
| Fred Sommers | <i>On Concepts of Truth in Natural Languages</i> |
| P. T. Geach | <i>The Perils of Pauline</i> |

CRITICAL STUDIES

- | | |
|------------------------|---|
| Frederick C. Copleston | <i>The Encyclopedia of Philosophy</i> |
| Louis Mackey | <i>Philosophy and Poetry in Kierkegaard</i> |

DISCUSSION

- | | |
|-----------------|---------------------------|
| Charles H. Kahn | <i>More on Parmenides</i> |
|-----------------|---------------------------|

BOOKS RECEIVED

- | | |
|------------------------------------|-------------------------------|
| Stanley O. Hoerr, Jr.
and Staff | <i>Summaries and Comments</i> |
|------------------------------------|-------------------------------|

PHILOSOPHICAL ABSTRACTS

ANNOUNCEMENTS

Annual Subscription \$6.00 Student Subscription \$4.00

Lyman-Beecher Hall, Haverford College, Haverford, Pa. 19041 USA

TWENTY-YEAR CUMULATIVE INDEX,
1947—1967 is now available for \$3.00 a copy

AMERICAN PHILOSOPHICAL QUARTERLY

MONOGRAPH SERIES

Edited by NICHOLAS RESCHER

The *American Philosophical Quarterly* also publishes a supplementary Monograph Series. Apart from the publication of original scholarly monographs, this is to include occasional collections of original articles on a common theme. The current monograph is included in the subscription, and past monographs can be purchased separately but are available to individual subscribers at *half price* (though not to institutional subscribers).

No. 1. STUDIES IN MORAL PHILOSOPHY. *Contents:* Kai Nielsen, "On Moral Truth"; Jesse Kalin, "On Ethical Egoism"; G. P. Henderson, "Moral Nihilism"; Michael Stocker, "Supererogation and Duties"; Lawrence Haworth, "Utility and Rights"; David Braybrooke, "Let Needs Diminish That Preferences May Prosper"; and Jerome B. Schneewind, "Whewell's Ethics." 1968, \$6.00.

No. 2. STUDIES IN LOGICAL THEORY. *Contents:* Montgomery Furth, "Two Types of Denotation"; Jaakko Hintikka, "Language-Games for Quantifiers"; James W. Cornman, "Types, Categories, and Nonsense"; Robert C. Stalnaker, "A Theory of Conditionals"; Alan Hausman and Charles Echelbarger, "Goodman's Nominalism"; Ted Honderich, "Truth: Austin, Strawson, Warnock"; and Colwyn Williamson, "Propositions and Abstract Propositions." 1968, \$6.00.

No. 3. STUDIES IN THE PHILOSOPHY OF SCIENCE. *Contents:* Peter Achinstein, "Explanation"; Keith Lehrer, "Theoretical Terms and Inductive Inference"; Lawrence Sklar, "The Conventionality of Geometry"; Mario Bunge, "What Are Physical Theories?"; B. R. Grunstra, "The Plausibility of the Entrenchment Concept"; Simon Blackburn, "Goodman's Paradox"; Stephen Spielman, "Assuming, Ascertaining, and Inductive Probability"; Joseph Agassi, "Popper on Learning from Experience"; D. H. Mellor, "Physics and Furniture"; and Michael Slote, "Religion, Science, and the Extraordinary." 1969, \$6.00.

AMERICAN PHILOSOPHICAL QUARTERLY

Edited by

NICHOLAS RESCHER

With the advice and assistance of the Board of Editorial Consultants:

Virgil C. Aldrich	James M. Edie	Benson Mates
Alan R. Anderson	José Ferrater-Mora	John A. Passmore
Kurt Baier	Richard M. Gale	Richard H. Popkin
Stephen F. Barker	Peter Thomas Geach	Richard Rorty
Monroe Beardsley	Adolf Grünbaum	George A. Schrader
Nuel D. Belnap, Jr.	Carl G. Hempel	Michael Scriven
Roderick M. Chisholm	John Hospers	Wilfrid Sellars
L. Jonathan Cohen	Raymond Klubansky	Alexander Sesonske
James Collins	Hughes Leblanc	Manley H. Thompson, Jr.
Arthur C. Danto	Ernan McMullin	John W. Yolton

Volume 7/Number 2

APRIL, 1970

CONTENTS

I. ALEX C. MICHALOS: <i>Decision-making in Committees</i>	91	V. FELIX E. OPPENHEIM: <i>Egalitarianism as a Descriptive Concept</i>	143
II. ALAN GEWIRTH: <i>Must One Play the Moral Language Game?</i>	107	VI. BRIAN SKYRMS: <i>Return of the Liar: Three-valued Logic and the Concept of Truth</i>	153
III. CLEMENT DORÉ: <i>An Examination of the "Soul Making" Theodicy</i>	119	VII. DAVID H. SANFORD: <i>Disjunctive Predicates</i>	162
IV. FRANCIS E. SPARSHOTT: <i>Disputed Evaluations</i>	131	<i>Books Received</i>	171

AMERICAN PHILOSOPHICAL QUARTERLY

POLICY

The *American Philosophical Quarterly* welcomes articles by philosophers of any country on any aspect of philosophy, substantive or historical. However, only self-sufficient articles will be published, and not news items, book reviews, critical notices, or "discussion notes."

MANUSCRIPTS

Contributions may be as short as 2,000 words or as long as 25,000. All manuscripts should be typewritten with wide margins, and at least double spacing between the lines. Footnotes should be used sparingly and should be numbered consecutively. They should also be typed with wide margins and double spacing. The original copy, not a carbon, should be submitted; authors should always retain at least one copy of their articles.

COMMUNICATIONS

Articles for publication, and all other editorial communications and enquiries, should be addressed to: The Editor, *American Philosophical Quarterly*, Department of Philosophy, University of Pittsburgh, Pittsburgh, Pennsylvania 15213.

OFFPRINTS

Authors will receive gratis two copies of the issue containing their contribution. Offprints can be purchased through arrangements made when checking proof. They will be charged for as follows: The first 50 offprints of 4 pages (or fraction thereof) cost \$12, increasing by \$1 for each additional 4 pages. Additional groups of 50 offprints of 4 pages cost \$8, increasing by \$1 for each additional 4 pages. Covers will be provided for offprints at a cost of \$4 per group of 50.

SUBSCRIPTIONS

The price *per annum* is eight dollars for individual subscribers and fourteen dollars for institutions. Checks and money orders should be made payable to the *American Philosophical Quarterly*. All back issues are available and are sold at the rate of three dollars to individuals, and four dollars to institutions. All correspondence regarding subscriptions and back orders should be addressed directly to the publisher (Basil Blackwell, Broad Street, Oxford, England).

MONOGRAPH SERIES

The *American Philosophical Quarterly* also publishes a supplementary Monograph Series. Apart from the publication of original scholarly monographs, this includes occasional collections of original articles on a common theme. The current monograph is included in the subscription, and past monographs can be purchased separately but are available to individual subscribers at a substantially reduced price. The back cover of the journal may be consulted for details.



I. DECISION-MAKING IN COMMITTEES

ALEX C. MICHALOS

I. INTRODUCTION

THIS paper has two aims, one fairly general and the other fairly specific. The general aim is to provide a basis for logical and empirical investigations of a particular type of group-decision procedure. The type is distinguished by the fact that inequalities in the "weight" of individual voters and their votes are permitted. For reasons which will become clear shortly, if such procedures are practicable at all, they would seem to be so only for small groups, say, of two to fifteen people. Since most committees contain less than fifteen members, if such procedures are practicable at all, they should be practicable for committees. Hence, our discussion is oriented toward decision-making in committees.

The more specific aim of this paper is to introduce and critically evaluate certain *prima facie* plausible solutions to the problem of selecting an acceptable group-decision procedure. We begin with a precise statement of the problem in Sect. II. In Sect. III five informal necessary conditions of adequacy for proposed solutions are introduced. Sect. IV introduces the notion of "weights of influence" and defines seven fundamental types of distributions. This is followed by a presentation of six more or less plausible "voting schemes" that involve the application of various "weights of influence." In Sects. VI and VII the constructive work of Sects. IV and V is examined in the light of our adequacy conditions. The upshot of this examination is, unfortunately, that none of the six schemes appears extraordinarily attractive.

II. THE PROBLEM

Very often individuals are called upon to share their decision-making capacities with others. They are asked to serve on committees. There are many reasons for establishing a committee, e.g., to keep people busy (Idle hands get into trouble!), to give

them a sense of "belonging," to keep them interested, to fix responsibility, to separate problems and solutions (divide and conquer), and so on. Some committees are supposed to solve problems or make decisions. They frequently plan conventions, parties, and dances, i.e., they *choose* locations, entertainment, refreshments. They *select* textbooks, candidates for new positions, visiting professors, office furniture, etc.

To avoid difficult epistemological problems which for present purposes need not be solved, we shall assume that a *decision is made* if and only if a person or group of people come to accept a certain sentence. The sentence would normally be in the imperative or in the indicative mode. Since committees are social organizations with their own rules, norms, procedures, etc., ordinarily a *committee makes a decision* if and only if there is a sentence which is accepted by the committee according to its operating rules. The operating rules of many committees are such that when there are conflicts of interest, the decision of the committee might not be similar to those of any of its members, e.g., a "compromise candidate" may become the first choice of a committee in which no member ranked him first.

Given the sense of "a committee decision" explained in the last paragraph, the primary question we shall attempt to answer in this paper is: How *should* the various decisions (with respect to some issue) of the members of a committee be amalgamated for the committee decision to be determined in the most acceptable fashion?¹ Three points must be made immediately about this question. In the first place, although it is a normative sort of question, it cannot be answered by logical analysis and intuition alone. What we ought to do to reach more (rather than less) acceptable committee decisions depends, among other things, on the *composition* of and the *issues* considered by this or that committee. That is to say, proposals for optimal group decision-making must be guided by some knowledge of the behavior, skills, resources,

¹ The best introductory survey of recent work on amalgamation procedures is R. Duncan Luce and Howard Raiffa, *Games and Decisions* (New York, John Wiley and Sons, Inc., 1964), pp. 327-371.

preferences, ambitions, etc., of human beings in fairly well-defined circumstances. Such knowledge is not only usually meager, it is also usually difficult to obtain. In the second place, it should be noted that in this paper the term "acceptable" will be used in a very broad sense to designate characteristics or dispositions frequently referred to by such (equally vague) terms as "competent," "wise," "good," "skillful," "enlightened," "rational," and so on. For the purposes at hand this usage is innocuous and convenient. Finally, it should be emphasized that we are primarily concerned with the procedures, methods, or means used to arrive at committee decisions, *not* with the products, ends, or committee decisions themselves.² Although in practice a peculiar committee decision often leads us to question the procedure used to obtain it (just as we might have suspicions about an argument schema which yielded a bizarre conclusion from ostensibly true premisses), it is important to distinguish the former from the latter, the *decision* from the *decision procedure*.

III. CRITERIA OF ADEQUACY

Lest we become dizzy chasing our own tails, something should be said about the *necessary* conditions of adequacy that might be applied to the amalgamation procedures offered as solutions to our problem. The following five are consistent (jointly satisfiable) and seem fairly plausible, although I suspect that they are neither independent nor in any technical sense complete.³ As we shall see later, they are considerably stronger than they appear.

- (3.1) The procedures must be *free from coercion* in the sense that committeemen are not

forced to cast their votes contrary to their preferences. This is to guarantee each voter the opportunity to use his influence (the "weight" of his vote) exactly as he sees fit.

- (3.2) The procedures must be self-consistent or *logically coherent*. This is merely a formal requirement excluding inconsistent procedures from which contradictory decisions could be obtained.

- (3.3) The procedures must be *practicable* or manageable. Practicability is bound to be relative to people and issues, so this is bound to be a pretty unstable criterion. But if a procedure is so complicated that for every group of people and every sort of issue some *other* procedure is simpler and equally effective, then our procedure would have to be regarded as relatively (relative to the other procedure) worthless. Hence, only those procedures that are manageable for *most* (normal) people and *some* (realistic) issues will be accepted. If as a matter of fact *most issues* confronting most committees do not merit a more subtle decision procedure than, say, majority rule by a show of hands, then our proposal(s) might still be acceptable and valuable, i.e., we only require practicability (which I am assuming means *some kind of superiority*) for *some* fairly realistic issues.⁴

- (3.4) The procedures must be *efficient* in the sense that they take account of all of the relevant information available to a committee for a given decision at a given time. What is required is not an everlasting

² In Karl Mannheim's terms, our concern is with "means rationality" rather than "substantial rationality." *Man and Society in an Age of Reconstruction* (London, Routledge and Kegan Paul, 1940), pp. 51-53.

³ Readers who are familiar with the literature on "social-welfare functions" which arose out of Kenneth J. Arrow's classic *Social Choice and Individual Values* (New York, John Wiley and Sons, Inc., 1963) will notice that the five conditions introduced below differ from Arrow's. In part this is merely a result of the fact that Arrow's problem and primary concerns differ from mine. However, it is also partially the result of my rejection of some of his basic assumptions. Roughly speaking, the two sets of conditions are related as follows. My condition (3.1) implies his conditions of nondictatorship and nonimposition. Conditions (3.2)-(3.5) are logically independent of his conditions of collective rationality, pair-wise comparisons, and Pareto optimality. Collective rationality and pair-wise comparisons are explicitly rejected in Sect. VII (below) and I have nothing to say about Pareto optimality. Similarly, so far as I know, Arrow has been silent regarding conditions analogous to (3.3)-(3.4), would probably reject (3.5), and accept (3.2). The most thorough discussion of Arrow's conditions is in Jerome Rothenberg, *The Measurement of Social Welfare* (Englewood Cliffs, N.J., Prentice-Hall, Inc., 1961).

⁴ There is some evidence that "group decision-making" in political parties through *prima facie* democratic processes is merely a "ritual" or "ceremonial" performing a legitimizing rather than a decision-making service; e.g., in Robert A. Dahl, *Who Governs?* (New Haven, Conn., Yale University Press, 1966), pp. 112-114. My impression is that most of the university committees I have served on have been providing the same (perhaps important and necessary) service, a service that would not seem to require a very subtle procedure.

search for relevant data, but the elimination of procedures that cannot process or make use of relevant data already possessed by committeemen.⁵

- (3.5) The *influence* of each committeeman should be *proportionate* to his relative competence. This condition requires a more lengthy explanation and some defense, and I shall begin with the former.

We suppose a committee decision is to be reached by "combining" the decisions of all of its members. If the committee operates on democratic principles, each member's decision influences the final outcome (the committee decision) in exactly the same way. Each member receives one vote and if it is cast at all, it carries as much weight or has as much influence *formally* (or *legally*, or *arithmetically*) on the final outcome as every other vote. (Informally or in fact, of course, every group seems to have leaders and followers, and both types of people influence each other in more or less subtle ways.) The committee decision itself then is usually determined by the majority of the voters. (3.5) implies that this democratic procedure is acceptable if and only if every voter is exactly as competent as every other. According to (3.5), each member's vote should carry only as much weight or have as much influence on the committee decision as his competence merits. Moreover, since the composition of and the issues confronted by committees varies, (3.5) suggests that the status of a committeeman's competence should be regarded as *relative* to both of these variables.

In view of the vast amount of literature devoted to the problem of equality,⁶ my defense of (3.5) will probably seem much too brief and primitive. It is as follows. As the influence of a committeeman increases, the chances of committee decisions being

similar to his increase (provided, of course, that other things are equal and stable). As the competence (ability, rationality, etc.) of a committeeman increases, the chances of his decisions being "correct" increase. Hence, as the influence and competence of a committeeman increase, the chances of committee decisions being "correct" increase. Similarly, it is easy to see that as the influence and incompetence of a committeeman increase, the chances of committee decisions being "incorrect" increase. According to (3.5), acceptable procedures must contain some provision for the distribution of influence according to relative competence. Therefore, insofar as this condition is met, the chances of committee decisions being "correct" will be increased and the chances of them being "incorrect" will be decreased. While this is not a guarantee that a committee will make "correct" decisions (or even *more* "correct" decisions *more* often), it does seem to be a necessary condition of adequacy for an acceptable group-decision procedure.⁷

Notice that the argument we have just presented does not commit us to the view that in all, most, or even some situations there is a "correct," more or less "correct," etc., decision which is *known* or even, practically speaking, *could be known* by certain committeemen. Of course in certain fields (e.g., logic and mathematics) there are independent and explicit criteria of "correctness" for the solutions of most problems. So in these fields one frequently can and does *know* whether or not his decision is "correct." However, in many of the most interesting and important fields (e.g., human welfare, morality, law, religion, etc.) one seldom has such knowledge. But it does not follow that there are no "correct," "incorrect," etc., decisions to be made with respect to the various sorts of evaluative problems that arise in these fields. Indeed, following a long line of different types of "objectivists" and a

⁵ I include this remark for those who might expect this condition to generate problems similar to those of Carnap's inductive "principle of total evidence"; e.g., A. J. Ayer, "The Conception of Probability as a Logical Relation," in S. Körner (ed.), *Observation and Interpretation in the Philosophy of Physics* (New York, Dover Publications, Inc., 1957), pp. 14-17.

⁶ A review of this literature may be found in George L. Abernethy (ed.), *The Idea of Equality* (Richmond, Va., John Knox Press, 1959).

⁷ Cf., "To have no voice in what are partly his own concerns is a thing which nobody willingly submits to; but when what is partly his concern is also partly another's, and he feels the other to understand the subject better than himself, that the other's opinion should be counted for more than his own accords with his expectations. . . ." John Stuart Mill, *Considerations on Representative Government* (New York, Harper and Brothers, 1867), p. 181. More recently James W. Prothro and Charles M. Grigg found that while roughly 96 percent of the people in their survey (of registered voters in Ann Arbor, Michigan and Tallahassee, Florida) agreed that "every citizen should have an equal chance to influence government policy," there was nearly a "total absence of consensus" on the view that "in a city referendum, only people who are well informed about the problem being voted on should be allowed to vote." "Fundamental Principles of Democracy," *The Journal of Politics*, vol. 22 (1960), pp. 282-285.

somewhat shorter line of naturalists,⁸ I believe that in one way or another evaluative problems (including problems of conflict of interest, etc.) can be analyzed in ways which permit one to *make* decisions that are "correct," more or less "correct," etc., and occasionally to *know* which are which. If it should turn out that a naturalistic theory of value is untenable, then it might well be necessary to make a sharp division between procedures that are applicable to evaluative problems and those that are applicable to nonevaluative problems. In that case, the argument presented above would be applicable only to procedures designed for nonevaluative problems. Nevertheless, until "all the returns are in," I shall continue to line up with the naturalists.

IV. WEIGHTING VOTERS

Even if one is persuaded by this argument for (3.5), which I am, it does not take us very far. For it is one thing to say that we *ought* to distribute influence according to relative competence and it is quite another to *know how* to do it.⁹ The remaining paragraphs of this section will be devoted to the latter problem, i.e., the problem of implementation.

Suppose we have a committee with n voting members:

$$M_1, M_2, \dots, M_n$$

Each member's vote will be assigned a *weight of influence* (or, for short, a *weight*) W which satisfies the following condition:

- (4.1) The weight of every committeeman's vote will be a real number greater than or equal to zero, i.e.,
 $W_i \geq 0$ ($i = 1, 2, \dots, n$)

In view of (3.5) it is imperative that the weight of every member's vote is proportionate to his relative competence. The crucial question is: How

are we going to measure relative competence? Clearly the usefulness of our amalgamation proposals will be severely curtailed unless a fairly plausible answer to this question is produced. The solution offered here is not new¹⁰ and it is *at best* fairly plausible. Its greatest virtue is its apparent ease of application, i.e., it is relatively convenient.¹¹ We shall explain it with the help of a concrete example.

Consider a committee with three members:

$$M_1, M_2, M_3$$

The committee is supposed to select an appropriate logic text for use in a freshman course. The usual (democratic) procedure is to give each committeeman one vote which he may cast as he sees fit. He may vote for this or that text, or abstain altogether. Following this democratic tradition, each of our committeemen will also be given exactly one vote. However, instead of asking each member to cast his vote on some question before the committee (e.g., the selection of a logic text), we shall ask each of them to *distribute* the weight of his vote (which is unity) among each of the committeemen. The distribution of a given member's vote should be such that it indicates the relative competence of every voter in the judgment of that committeeman.

More precisely, we are asking *first* that each member *weakly order* every voter in accordance with his relative competence. That is, every committeeman is supposed to try to decide for all of the members whether

- (i) M_i is more competent than M_k or M_k is more competent than M_j , or M_i and M_k are equally competent,

and

- (ii) if M_i is at least as competent as M_k and M_k is at least as competent as M_j , then M_i is at least as competent as M_j .

⁸ This shorter line includes such illustrious names as M. R. Cohen, John Dewey, Herbert Feigl, R. W. Sellars, and most recently, Henry Margenau. The latter's *Ethics and Science* (Princeton, N.J., D. Van Nostrand Company, 1964) may be regarded as an outline of the sort of view I have adopted.

⁹ Mill's suggestion for "weighting" voters in a "universal but graduated suffrage" included the administration of "a trustworthy system of general examination," consideration of "the nature of a person's occupation," "successful performance" of a trade, "graduates of universities," and excluded "sex" or "any pecuniary qualification." *Op. cit.*, pp. 182-192. Some writers have admitted unequal "weightings" into their formal theories without tackling the problem of how such "weights" could be obtained; e.g., Germain Kreweras, "A Model to Weight Individual Authority in a Group," in S. Sternberg, V. Capeccchi, T. Kloek, and C. T. Leenders (eds.), *Mathematics and Social Sciences I* (Paris, Mouton and Co., 1965), pp. 111-118; and William H. Riker, *The Theory of Political Coalitions* (New Haven, Conn., Yale University Press, 1962), pp. 257-258.

¹⁰ Roughly the same procedure was used by J. Sayer Minas and Russell L. Ackoff in "Individual and Collective Value Judgments," in Maynard W. Shelly, II and Glenn L. Bryan (eds.), *Human Judgments and Optimality* (New York, John Wiley and Sons, Inc., 1954), pp. 351-359. A considerably more complicated (though not necessarily more reliable) nine-step procedure may be found in my unpublished paper "Split Weight Option Voting."

¹¹ A number of doubts about its "greatest virtue" are raised in Sect. VII.

And *second* we are asking that every member try to assign a number to every voter in accordance with his rank order to serve as an *indicant*¹² of his relative competence. These numbers will be called "initial weights of influence" (or, "initial weights"). Moreover,

- (4.2) Every initial weight will be a real number w in the closed interval from 0 to 1, i.e.,

$$0 \leq w \leq 1$$

and the sum of the assignments that may be given by every committeeman will be 1.

The following matrix illustrates a possible result of (4.2).

FIGURE I

	M_1	M_2	M_3	w_i
M_1	1	0	0	1
M_2	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	1
M_3	$\frac{1}{2}$	$\frac{1}{2}$	0	1
W_i	$1\frac{1}{6}$	$\frac{5}{6}$	$\frac{1}{3}$	$3 = n$

Here M_1 has assigned himself an initial weight of 1 and left nothing for M_2 or M_3 . We may assume then, that in the opinion of M_1 the other members of the committee do not know anything about logic texts. Or, to put the point another way, we may assume that M_1 believes the committee decision has a better chance of being "right" if it is similar to his own decision. M_2 believes his opinion is worth no more and no less than the others. So he distributes his initial weight equally among all the voters. M_3 figures that M_1 knows about as much as M_2 about logic texts and that he (M_3) just does not know anything. So he distributes his initial weight equally between M_1 and M_2 .

If a committee has n members then it has n votes to be distributed or, possibly, withheld. Then, by definition:

- (4.3) The weight of every committeeman's vote equals the sum of the initial weights assigned to him by all members of the committee, i.e.,

$$W_j = \sum w_{ji} \quad (i \text{ and } j = 1, 2, \dots, n)$$

for the j th committeeman according to each (i th) member.

In Figure I, the weight of M_1 equals the sum of the initial weights in the column below M_1 , namely, $1\frac{1}{6}$. The weights of M_2 and M_3 are $\frac{5}{6}$ and $\frac{1}{3}$, respectively. The sum of these weights is $n = 3$, which happens to equal the maximum number possible for this group because no one withheld any of his initial weight.¹³

Notice the efficiency of the voting procedure prescribed for these committeemen in comparison to the usual (democratic) procedure. For example, consider M_3 's position. Given the usual procedure, if M_1 and M_2 disagreed on a text then M_3 would have to *either* abstain altogether *or else* go along with M_1 or M_2 . But in *both* cases he would be merely making the best of a bad situation. His "real" judgment is that M_1 's views are about as reliable as M_2 's. However, the voting procedure forces him to act as if he had nothing at all to contribute (i.e., he abstains) or else to act as if he preferred the view of M_1 or M_2 . In short, the usual procedure suppresses that modicum of competence that the committee recognizes in M_3 . On the other hand, our procedure of weighting voters takes account of M_3 's ability. It allows him to make a contribution that reflects his own best judgment. It is *not* a judgment about the question at issue (i.e., about the best logic text), but about the relative ability of the other members of the committee to judge that issue. It is what many writers regard as a typically administrative decision.¹⁴ It is a decision about the ability of certain personnel to judge a certain issue, rather than a technical decision about the issue itself. The legitimacy and usefulness of administrative decisions can hardly be doubted, since it is difficult to imagine a highly complex organization such as an industrial corporation, university, or government agency without many people responsible for such decisions, viz., the administrators or managers. Therefore, if all other things are equal, then insofar as the procedure we are recommending provides a more efficient treatment of such

¹² "The difference between an *indicant* and a *measure* is just this: the indicant is a presumed effect or correlate bearing an unknown . . . relation to some underlying phenomenon, whereas a measure is a scaled value of the phenomenon itself. Indicants have the advantage of convenience. Measures have the advantage of validity. We aspire to measures, but we are often forced to settle for less." (Italics added.) S. S. Stevens, "Mathematics, Measurement and Psychophysics," *Handbook of Experimental Psychology*, ed. by S. S. Stevens (New York, John Wiley and Sons, Inc., 1951), p. 48.

¹³ To simplify the illustrations, no weight is withheld in any of them.

¹⁴ E.g., "In the field of organization, the knowledge on which what we call responsible control depends is not knowledge of situations and problems and of means for effecting changes, but is knowledge of other men's knowledge of these things . . . the crucial decision is the selection of men to make decisions." Frank H. Knight, *Risk, Uncertainty and Profit* (New York, Harper and Row, Inc., 1965), pp. 292-299.

decisions than the usual procedure, the former should be regarded as superior.

According to (4.1)–(4.3), the weight of each committeeman's vote must be a real number in the closed interval from 0 to n , i.e.:

$$0 \leq W_j \leq n \quad (j = 1, 2, \dots, n)$$

If every committeeman distributes his initial weight in equal amounts among all of the members of a committee then each one will receive an initial weight of $1/n$ from every member. Since there are n members, the result of such a distribution will be a *democratic weighting* with every committeeman's weight:

$$W_j = 1 \quad (j = 1, 2, \dots, n)$$

Formally, this result is indistinguishable from an *anarchic weighting* in which everyone keeps all of his initial weight. The different attitudes of democrats and anarchists toward legitimate government is reflected by the different paths leading to the equalitarian weights. The democrat assigns each committeeman the same weight and, therefore, is obliged to use numbers of supporters as a legitimizing criterion. The anarchist assigns only himself a weight and, therefore, is obliged to use his own preferences as a legitimizing criterion.

If every committeeman distributes his initial weight such that the weight of a single member, say, M_a is

$$W_a = n$$

the result will be a *delegatory weighting*. In effect the complete responsibility for the committee decision has been delegated to M_a . If the latter could be *born* into the position, there would be some justification for regarding this pattern as a *monarchic weighting*.

If the distribution of initial weights is such that the total weight of fewer than half of the voters is greater than half of the total weight of the whole committee, the result is an *oligarchic weighting*. More precisely, if there are m committeemen such that

$$\sum W_i > n/2 \quad (i = 1, 2, \dots, m)$$

although

$$m < n/2$$

the weighting is oligarchic. If the composition or membership of the set of m committeemen varies as the issues before the committee vary, it might be more appropriate to refer to the weighting

as *polyarchic*. In practice, many organizations that operate *formally* with democratic weightings, operate *informally* with polyarchic or delegatory weightings.¹⁵

Finally, if the distribution of initial weights is such that the weight of a single member, say, M_o is

$$W_o = 0$$

the result is a *disfranchising weighting*. Notice that no one can be disfranchised unless he chooses to be.

Obviously these weighting patterns are not mutually exclusive in pairs. While no pattern can be both oligarchic and democratic (or democratic and disfranchising), every delegated pattern must be oligarchic. Moreover, *a priori* there seems to be no good reason to suspect that the patterns defined here exhaust the interesting possibilities, i.e., these weighting patterns are probably not exhaustive.

More importantly, perhaps, it should be emphasized that it is *not* being claimed that the specification of the type of weighting pattern employed in a committee is *sufficient* to characterize the committee as, say, democratic. It is true, however, that the specification of the type of pattern is *necessary* for such a characterization.

V. WEIGHTED VOTING

Now that we have a procedure for weighting voters, the question is: How should the weights be used to reach committee decisions? This question can be neatly divided into the following two questions:

- (A) How should weights be applied by voters?
- (B) What proportion of the total weight (n) available should be required for a committee decision?

The latter question (B) has received considerable attention by legislators and philosophers, and we shall not attempt to improve upon traditional discussions here.¹⁶ Instead, we shall focus on the former (A) and try to present some options that have received little or no attention. In particular, we shall consider six voting schemes. For each of the schemes it will be assumed that the decision (item, solution, policy, candidate, etc.) which receives the greatest support or has the largest weight will be accepted as the committee decision. Roughly speaking, this assumption is equivalent to

¹⁵ A classic description of a polyarchic system may be found in Dahl, *op. cit.*

¹⁶ An excellent review and criticism of some of the alternatives may be found in C. L. Dodgson (Lewis Carroll), "A Discussion of the Various Methods of Procedure in Conducting Elections" reprinted in Duncan Black, *The Theory of Committees and Elections* (Cambridge, Cambridge University Press, 1963), pp. 214–222.

answering (B) with: a plurality or simple majority is required for a committee decision. This is not an especially profound or unproblematic answer, but it will be adequate for our purposes. Above all it should be remembered that (A) and (B) are quite different questions, and that success or failure with either does not imply success or failure with the other.

Question (A) is vague. It might be answered by such disparate remarks as: In accordance with their consciences; with caution; happily; all together; in bits and pieces; etc. The last two replies suggest the sense in which we are interested. To begin with, we might apply weights as follows:

Total weight scheme: Each committeeman puts the total weight of his vote on a single decision (candidate, option, etc.), or he withholds all of it.

For example, suppose the weights assigned to a five-membered committee are:

FIGURE II

	M_1	M_2	M_3	M_4	M_5
W_i	1	0.5	1.3	0.7	1.5

and there are three alternative decisions to consider

D_1 D_2 D_3

Then on the total weight scheme, the result of a vote might be:

FIGURE III

	D_1	D_2	D_3
M_1	0	1	0
M_2	0.5	0	0
M_3	1.3	0	0
M_4	0	0	0.7
M_5	0	1.5	0
Total	1.8	2.5	0.7

That is, M_1 puts all of his weight behind D_2 , M_2 puts all of his weight behind D_1 ; and so on. The result is that D_2 becomes the committee decision because it has the most support.

The total weight scheme is perfectly straightforward. Each man either applies all of his influence (i.e., the total weight of his vote) to a single decision or he withholds all of it. However, sometimes a committeeman regards certain alternative decisions as equally acceptable. Or, faced with three or more options, he is frequently able to weakly order them. If *any* of the members of a

committee are able to weakly order the various alternatives before them, then it would be useful (i.e., more efficient) to give them the option of dividing the weight of their votes according to their preferences. This involves two closely related assumptions. In the first place, it must be granted that weights of influence, which are already indicants of relative competence, may also be used as indicants of preferences. And secondly, it must be granted that people's preferences are comparable, e.g., if M_1 and M_2 rank D_1 above D_2 and the latter above D_3 , we would *grant* that they both "feel about the same" about the three decisions *instead of* insisting that, say, M_1 prefers D_1 to D_2 much more than M_2 prefers D_1 to D_2 , and so on.¹⁷ While the dual role of weights of influence seems harmless enough, we shall have more to say about the weak ordering and interpersonal comparisons of "utility" in Sect. VII.

Given the above assumptions, we might employ a:

Split weight option scheme: Each committeeman distributes the weight of his vote among the decisions in accordance with his preferences, or he withholds any part or all of it.

For example, consider the five-membered committee above, faced with the same three decisions. A possible result of the split weight option scheme for the voters in Figure II might be:

FIGURE IV

	D_1	D_2	D_3	W_i
M_1	0.5	0.5	0	1
M_2	0.5	0	0	0.5
M_3	0.8	0.5	0	1.3
M_4	0	0.2	0.5	0.7
M_5	0.5	1	0	1.5
Total	2.3	2.2	0.5	$5 = n$

Here M_1 distributes the weight of his vote between two equally attractive (acceptable, plausible, preferred, etc.) decisions D_1 and D_2 ; M_2 keeps his vote intact; and so on. The result is that D_1 narrowly becomes the committee decision.

A major advantage of the split weight option scheme over the total weight scheme is that the former *takes account* of discernible preferential differences while the latter cannot take account of such differences. Or, to put this important point in a slightly different way, a committee decision

¹⁷ A suggestive graphic representation of this problem is given by Nicholas Rescher, "Notes on Preference, Utility and Cost," *Synthese*, vol. 16 (1966), p. 333.

based on the split weight option scheme is a function of more subtle or precise judgments than a committee decision based on the total weight scheme. Hence, the former scheme is more efficient than the latter.

Instead of allowing committeemen to divide their weights to indicate their preferences, we might provide each voter with an additional unit to distribute. More precisely, we might let each voter assign every decision (item, policy, etc.) a real number U in the closed interval from 0 to 1 called a "utility," with the sum of the utilities assigned by each member equal to or less than one. Here the term "utility" means nothing more than the value (desirability, satisfactoriness, etc.) of a decision according to a committeeman. An example of the result of such assignments might be

FIGURE V

	D_1	D_2	D_3	sum
M_1	0.5	0.5	0	1
M_2	1	0	0	1
M_3	0.7	0.3	0	1
M_4	0	0.3	0.7	1
M_5	0.4	0.5	0.1	1

If U_{ij} is the utility of the i th decision according to the j th member then the total *weighted-utility* of that decision for that member is:

$$U_{ij} W_j$$

Now we may define the:

Total weighted-utility scheme: Each committeeman assigns every decision a total weighted-utility (i.e., the product of the total weight of influence of his vote and the utility he assigns to each decision).¹⁸

For example, given the utility assignments in Figure V and the weights of influence in Figure II, we have

FIGURE VI

	D_1	D_2	D_3
M_1	(1) (0.5)	(1) (0.5)	(1) (0)
M_2	(0.5) (1)	(0.5) (0)	(0.5) (0)
M_3	(1.3) (0.7)	(1.3) (0.3)	(1.3) (0)
M_4	(0.7) (0)	(0.7) (0.3)	(0.7) (0.7)
M_5	(1.5) (0.4)	(1.5) (0.5)	(1.5) (0.1)

Total weighted-utility sum	2.51	1.85	0.64
----------------------------	------	------	------

The result of applying the total weighted-utility scheme with the given numerical assignments is a clear victory for D_1 .

Although the split weight option scheme is less cumbersome than the total weighted-utility scheme, it is easy to show that the two schemes are equivalent. Using the former, a committeeman M_j would distribute the weight of his vote W_j such that every alternative D_i receives a certain proportion of it, say, x_{ij} where

$$\sum x_{ij} \leq W_j \quad (i = 1, 2, \dots, n)$$

for an n -membered committee. Using the latter, M_j would distribute his unit of utility U_j among the members in the same proportions. So, for any member M_j and any alternative D_i :

$$\frac{x_{ij}}{W_j} = \frac{U_{ij}}{1} = U_{ij}$$

Hence:

$$x_{ij} = U_{ij} W_j$$

But x_{ij} is the index of M_j 's support for D_i according to the split weight option scheme and $U_{ij} W_j$ is the same index according to the total weighted-utility scheme. Therefore, the two schemes are equivalent.

To obtain greater uniformity in our utility assignments, we might employ the sort of marking system introduced by Jean-Charles Borda.¹⁹ We would assign 0 utility to the least preferred alternative, 1 to the alternative immediately above that, 2 to the third highest, and so on until we reach the end of the lot. Equally preferred alternatives are assigned the same number. For convenience, we shall call such indicants "Borda-utilities." The preferences ranked in Figure V would be indicated as follows in terms of Borda-utilities.

FIGURE VII

	D_1	D_2	D_3
M_1	1	1	0
M_2	1	0	0
M_3	2	1	0
M_4	0	1	2
M_5	1	2	0

Notice that while alternatives with Borda-utilities of n have exactly n ranks below them, they may have more or less than n alternatives below them. For example, M_2 distinguishes two preference levels. So the top level has a Borda-utility of $n = 1$, and there is one level but two alternatives below D_1 .

¹⁸ This is basically the scheme recommended by Minas and Ackoff, *op. cit.*, pp. 355-356.

¹⁹ "Mémoire sur les élections au scrutin," *Histoire de l'Académie* (1781). For a thorough examination of Borda's views see Black, *op. cit.*, pp. 59-66, 156-159.

Now we may define a:

Weighted Borda-utility scheme: Each committeeman assigns every decision a weighted Borda-utility (i.e., the product of the total weight of influence of his vote and the Borda-utility he assigns to each decision).

For example, given the Borda-utility assignments of Figure VII and the weights of influence of Figure II, we have:

FIGURE VIII

	D_1		D_2		D_3	
M_1	(1)	(1)	(1)	(1)	(1)	(0)
M_2	(0.5)	(1)	(0.5)	(0)	(0.5)	(0)
M_3	(1.3)	(2)	(1.3)	(1)	(1.3)	(0)
M_4	(0.7)	(0)	(0.7)	(1)	(0.7)	(2)
M_5	(1.5)	(1)	(1.5)	(2)	(1.5)	(0)

Weighted Borda-utility sum	5.6	6	1.4
----------------------------	-----	---	-----

Hence, the result of applying the weighted Borda-utility scheme with the given numerical assignments is a victory for D_2 .

If we provide additional units such as utilities and Borda-utilities to be distributed and allow committeemen to split the weight of their votes, then we may define a:

Split weight utility scheme: Each committeeman assigns every decision a split weight utility (i.e., the product of some or all of the weight of influence of his vote and the utility he assigns to each decision).

and a:

Split weight Borda-utility scheme: Each committeeman assigns every decision a split weight Borda-utility (i.e., the product of some or all of the weight of influence of his vote and the Borda-utility he assigns to each decision).

Using the Borda-utility assignments of Figure VII and the weights of influence of Figure II, Figure IX illustrates a possible result of applying the latter scheme:

FIGURE IX

	D_1		D_2		D_3	
M_1	(1)	(1)	(0.5)	(1)	(0.5)	(0)
M_2	(0.5)	(1)	(0.2)	(0)	(0.5)	(0)
M_3	(1.3)	(2)	(0.3)	(1)	(0.8)	(0)
M_4	(0.7)	(0)	(0.7)	(1)	(0.7)	(2)
M_5	(1.5)	(1)	(1)	(2)	(0.5)	(0)

Split weight Borda-utility sum	5.6	3.5	1.4
--------------------------------	-----	-----	-----

Hence, the result of applying this scheme with the given numerical assignments is a victory for D_2 . (Notice that every committeeman is permitted to multiply the Borda-utility he assigns to each decision by a number no greater than his total weight.)

By a line of argumentation analogous to that used to show that the split weight option and total weighted-utility schemes are equivalent, it may be shown that the split weight utility and split weight Borda-utility schemes are both equivalent to the former two schemes. Thus, in view of the unnecessary labor involved in the additional computations required for the application of the latter three in comparison with the split weight option scheme, there is no doubt that the other three schemes will never be practicable.

The following matrix summarizes the six schemes we have considered:

	No Additional Units Provided	Additional Units Provided
Use All or No Weight	total weight	{ total weighted-utility weighted Borda-utility
Use Some, All, or No Weight	split weight option	{ split weight utility split weight Borda-utility

The total weight scheme comes about by answering question (A), (How should weights be applied by voters?) with "as a whole" and not providing any additional units to be distributed. If (A) is answered by "as a whole or in parts" and no additional units are provided, the result is the split weight option scheme. The remaining entries in the matrix arise analogously.

VI. OBJECTIONS AND REPLIES: WEIGHTING VOTERS

The constructive work of this essay has now been completed, and it is time to appraise our results in the light of the adequacy criteria introduced in Sect. III. Six solutions to our basic problem (viz., How should the decisions [with respect to some issue] of the members of a committee be amalgamated for the committee decision to be determined in the most acceptable fashion?) were suggested in Sect. V. Each of these six employed the procedure for weighting voters that was described in Sect. IV. So, if there are no acceptable procedures for weighting voters, then the practical

value of the six schemes of weighted voting must be nil. It will be convenient then, to consider objections to the procedure for weighting voters prior to and in isolation from objections to particular schemes of weighted voting.

How does our procedure for weighting voters fare with respect to our five criteria, viz., (3.1) freedom from coercion, (3.2) internal consistency, (3.3) practicability, (3.4) efficiency, and (3.5) influence proportionate to relative competence? Taking the least problematic first, our procedure seems (to me) to satisfy the conditions of freedom from coercion, internal consistency, and efficiency. But the other two requirements require some discussion.

To begin with, it might be objected that our procedure violates (3.5) because it begins with an equalitarian distribution of influence in the form of a single vote worth exactly one unit. That is, *before* anyone's competence is estimated, every committeeman is given a vote whose initial weight is equal to that of every other. But according to (3.5), such a distribution is permissible only if the competence of every committeeman is equal to that of every other. However, at this (first) stage in our procedure we do not know how competent any committeeman is. Hence, we cannot justify an opening equalitarian distribution. Moreover, if we *could* justify such a distribution then the whole procedure of weighting voters would be a waste of time, because they would all receive the same weight always. But the first step in our procedure must be either unjustified or justified. Therefore, it must either violate (3.5) or trivialize the whole procedure.

This dilemma is more apparent than real. The first horn is problematic but the second horn is based on a misunderstanding. Let us take the second horn first. Recall that in our procedure each committeeman is called upon to make two different *types* of decisions, one on the abilities of the other members and the other on the issue(s) before the committee. Our opening equalitarian distribution only presupposes that every committeeman is equally qualified to make the first type of judgment. If *this* assumption is warranted then all that follows is that every member's *opinion about every committeeman's judgment* about the issue(s) before the committee is equal to every other member's opinion. It does *not* follow that every member's *opinion about the issue(s)* before the committee is

equal to every other member's opinion. Hence, the whole weighting procedure cannot be trivialized by an opening equalitarian distribution. So much for the second horn.

The first horn of the dilemma cannot be dispatched so swiftly. Indeed, we must grant immediately that we have no reason to suppose that the ability to judge the ability of personnel to judge more or less technical issues is equally distributed among all or most people. What we would like to substitute for an opening equalitarian distribution is a distribution based on some reliable independent test. But this raises the question: How should we select such a test? And this question seems to lead us straight to Sect. II of this paper (which has finally led us to the present point) *or* to an infinite regress of reliable independent tests of reliable independent tests *or* to some more or less arbitrarily final reliable independent test. No doubt the last alternative will be preferred by most people, and the haggling will be over the cut-off point. The opening equalitarian distribution suggested in our procedure is merely a convenient (admittedly early) cut-off point which should be adequate for most purposes. After all, even the usual democratic procedure (which is less efficient than any of ours) has been regarded as adequate for most purposes. If the *cost* of an erroneous committee decision is very high, we would expect much more elaborate precautions to be taken to increase the chances of arriving at a more acceptable opening distribution.²⁰

Briefly, then, my reply to the first horn of the dilemma is as follows. The objection is sound but not totally destructive. It reveals a defect in our procedure which becomes more or less serious as the cost of errors increases or decreases, respectively, and which further research should attempt to eliminate or reduce as much as possible.

Supposing we had solutions to these problems, would our procedure then be practicable [condition (3.3)]? It seems that our procedure may be impracticable for *at least* eight reasons.

(1) Committeemen may find the task of weakly ordering their colleagues in accordance with relative competence too *complicated*. Competence is very likely multi-dimensional and the dimensions may be extremely difficult if not impossible to compare. For example, a certain committeeman may be very good at identifying, say, the pedagogical strengths and weaknesses of a textbook (e.g., the organization

²⁰ The relation between costs and precision in measurement is outlined in C. West Churchman, *Prediction and Optimal Decision* (Englewood Cliffs, N.J., Prentice-Hall, Inc., 1964), pp. 116ff.

and presentation of new material, definitions, illustrations, exercises, etc.) but very bad at recognizing factual or formal errors (e.g., that this was discovered by so-and-so then, that most people call this a such-and-such, that it is misleading to refer to this as that, etc.). Should such a committeeman be assigned a high or a low weight? Should a committeeman who is good at *both* tasks be assigned a weight that is *twice* as great as that of a committeeman who is only good at one? Are the two skills "really" equally important? Even if these problems could be solved, the cost of solving them may be too high *given* the benefits obtainable from our procedure *and* the availability of other procedures. That is, it is not merely this or that complication in the abstract that may be objectionable. Rather, it is the fact that in comparison with other procedures, the cost of untangling the complications in ours may be too high and the likely rewards too low to vindicate its application.

(2) The determination of relative competence may produce *disputes*, *factions*, and *enemies* that might otherwise be avoided. For example, a member may not appreciate receiving a low weight from a voter to whom the former has just assigned a low weight. Since committee meetings will probably take more time with our procedure, we may expect greater fatigue, irritability, impatience, and ennui. Hence, even trivial issues may become highly problematic.

(3) Our procedure may be regarded as too "*brutal*" because it does not provide any face-saving safeguards for perennial "light-weights." Even if the weights are never made public (which merely requires the help of a discreet and honest assistant, or a machine), it is bound to occur to some people that in their *own* opinion the judgment of certain committeemen is almost always nearly worthless. What is worse, if the split weight option scheme is used and everyone must be informed of their colleagues' views about their relative competence, then perennial "light-weights" are

going to be repeatedly embarrassed and probably embittered.

(4) Our procedure may breed *irresponsibility* in those who, for a certain issue, have been assigned a fairly low weight. The low weight assigned to a committeeman may have the effect of a self-fulfilling prediction, i.e., everyone expects him to contribute very little; so he expends very little effort and, consequently, he has very little to contribute.

(5) Committeemen may nullify the opportunity to determine the relative competence of their colleagues by employing a *minimax loss* strategy. According to this strategy, a voter should distribute his initial weight to insure himself the smallest possible loss. His options are to give away none of it, give away part of it, or give away all of it. Clearly, he can protect himself against any loss at all by simply keeping his vote all to himself.²¹ However, if everyone employs this strategy, then the result is an anarchic weighting. And if enough people employ this strategy often enough, our procedure would have to be regarded as redundant.²²

(6) It is likely that committeemen will *misjudge* the relative competence of their colleagues. They will commit "errors of leniency" (i.e., assign more weight to those they like), "central tendency" (i.e., avoid assigning extreme weights of 0 or 1), "logic" (i.e., assign similar weights for skills that seem to be "logically related"), "contrast" (i.e., assign weights roughly inversely proportional to those they assign themselves), and "proximity" (i.e., assign similar weights for skills that are considered at roughly the same time or in the same context).²³ They will fall prey to the "halo effect" (i.e., assign weights in accordance with their "general impression" of the individual rated)²⁴ and the "Matthew effect" (i.e., assign higher weights to the "biggest names" regardless of their particular competence in a given situation).²⁵ They will be influenced by "prestige considerations,"²⁶ and, as the ambiguity of the situation increases, by social pressure to

²¹ Similarly, it has been argued that "A second vital reason for seeking the condition of political equality is a strategic calculation. Like so many important decisions, this one, too, is a calculated risk. Reduced to its boldest terms, our strategy comes to this: we cannot be confident of continued membership in an elite group whose preferences would be counted for more rather than less." Robert A. Dahl and Charles E. Lindblom, *Politics, Economics and Welfare* (New York, Harper and Row, Inc., 1963), p. 43.

²² Other possible strategies may be found in my *Principles of Logic* (Englewood Cliffs, N.J., Prentice-Hall, Inc., 1969), ch. 8.

²³ Detailed discussions of these errors may be found in J. P. Guilford, *Psychometric Methods* (New York, McGraw-Hill Book Co., 1954), pp. 278-280.

²⁴ *Ibid.*, p. 279.

²⁵ Robert K. Merton, "The Matthew Effect in Science," *Science*, vol. 159 (1968), pp. 56-63.

²⁶ Robert E. Lane, "Political Personality and Electoral Choice," *American Political Science Review*, vol. 49 (1955), pp. 173-174.

"conform."²⁷ Hence, instead of indicating relative competence, our weights will be meaningless numbers misleadingly suggesting precision and accuracy.

(7) There is some evidence that less competent people are less influential than more competent people anyhow.²⁸ So an attempt *formally* to weaken the former and strengthen the latter is unnecessary.

(8) If a committeeman believes that some other member of a committee is especially competent, then, rather than giving the latter some of his weight (losing formal influence and flexibility certainly and informal influence possibly), he would be wise to keep all of his weight and simply allow himself to be guided by the other member's decision. After all, from the former's point of view, the final result of giving the latter his weight to put behind a certain decision is no different from putting it in the same place under the other member's supervision. Hence, because committeemen have something to lose by distributing their weight and nothing to lose by keeping it, few or no distributions should ever occur.

Let us consider each of these objections in turn.

(1) It must be granted that *some* people may find the task of weakly ordering their colleagues in accordance with relative competence virtually impossible. That is why our procedure does not *require* such an ordering from anyone. The ordering is *requested*, and if most people cannot satisfy the request, most of the time, then our procedure cannot be expected to accomplish anything most of the time. Nevertheless, our procedure provides an opportunity to use whatever resources happen to be available, which seems to make it superior to procedures without this provision. Moreover, and fortunately, there is no evidence of a widespread human inability to perform such tasks. On the

contrary, while it is difficult to appraise the relative competence of this or that scholar in a certain area, such appraisals are very common. Thousands of employers are required to do just that, and they are frequently promoted for doing it well and discharged for doing it poorly. Hence, the real question seems to be: Are the rewards of untangling the complications in our procedure high enough to justify the costs? And in the absence of a definite issue, committee, and alternative procedure this question seems to be unanswerable.

(2) There seem to be good reasons for believing that our procedure will not create an inordinate number of disputes. In the first place, people frequently judge the merits of candidates for various positions of leadership, and our procedure is merely an extension of this common practice. In the second place, certain precautions may be taken to reduce the chances of disputes. For instance, one should avoid using our procedure in situations in which the likelihood of misunderstanding is known to be high, e.g., in meetings involving people with quite specific and, perhaps, antagonistic *roles* such as labor and management representatives,²⁹ or officers and enlisted men.³⁰ Perhaps committeemen might be impressed with the idea that they are all part of a single "team" and that the less energy they expend attacking each other the more they will be able to expend on the issue before the committee. Again, the reward and punishment structure might be designed to militate against such phenomena, i.e., let the "payoffs" go to those who do not get involved in disputes and factions, and the penalties go to those who do.³¹

(3) The third objection seems to suggest a nice problem of balancing the demands of morality against those of rationality. Fortunately, that is not the case. The problem is one of balancing the

²⁷ E. L. Walker and R. W. Heyns, *An Anatomy for Conformity* (Englewood Cliffs, N.J., Prentice-Hall, Inc., 1962), pp. 92-95. Further problems may be found in Bernard R. Berelson, Paul F. Lazarsfeld, and William N. McPhee, *Voting* (Chicago, University of Chicago Press, 1962), p. 232, and Paul F. Lazarsfeld, Bernard Berelson, and Hazel Gaudet, *The People's Choice* (Ithaca, N.Y., Cornell University Press, 1960), pp. 80ff.

²⁸ E.g., "The structure of interpersonal relations within legislative systems functions to place legislators with high degrees of skill and conscientiousness in the center of the legislative system and to increase their influence, and to isolate legislators with low degrees of skill and conscientiousness on the periphery of the legislative system and to decrease their influence." Stephen V. Monsma, "Interpersonal Relations in the Legislative System: A Study of the 1964 Michigan House of Representatives," *Midwest Journal of Political Science*, vol. 10 (1966), p. 363. This article also contains a long bibliography. An instructive criticism of earlier attempts at formulating indices of "influence" may be found in Ralph K. Huitt, "The Outsider in the Senate," *American Political Science Review*, vol. 55 (1961), pp. 566-575.

²⁹ Mason Haire found that "management and labor each sees the other as less dependable than himself . . . and . . . deficient in thinking, emotional characteristics, and interpersonal relations . . ." in "Role-perception in Labor-management Relations: An Experimental Approach," *Industrial and Labor Relations Review*, vol. 8 (1955), p. 215.

³⁰ Paul F. Lazarsfeld, "The American Soldier," *Public Opinion Quarterly*, vol. 13 (1949), pp. 383-386.

³¹ A number of other techniques are described in Richard E. Walton and Robert B. McKersie, "Behavioral Dilemmas in Mixed-motive Decision Making," *Behavioral Science*, vol. 11 (1966), pp. 381-384.

demands of one moral prescription against those of another. On the one hand, committeemen are morally obliged to try to avoid embarrassing their colleagues. On the other, they are morally obliged to try to make as "correct" a decision as possible. Hence, the question they must ask themselves is this: Is the moral cost of embarrassing (and, perhaps, embittering) some members greater than, equal to, or less than the moral cost of making a less "correct" decision. If it costs more (i.e., seems to create more feelings of guilt) to identify and isolate the "light-weights" than it does to accept a slightly "incorrect" committee decision, then a committeeman ought to accept the latter; otherwise he ought to choose the former, unless the costs are equally balanced. In the latter case he might flip a coin.

It would be a mistake to assume that "light-weights" *must* be embarrassed or embittered. While there is a wealth of evidence to support the view that they will probably respond that way,³² there is also evidence that the reaction of individuals to participation in or exclusion from decision-making that affects their lives is significantly influenced by their *expectations*.³³ Hence, someone who expects "the other fellow" to be responsible for this or that may be perfectly happy to be assigned a lighter weight. Moreover, it should be noted again that if the variety of issues considered by a committee is large, the probability that a certain committeeman will always be judged a "light-weight" is small. Most people tend to be specialists by inclination and aptitude. Hence, if the issues before a committee *vary*, it is likely that no one will be a perennial "light-weight" or "heavy-weight," i.e., the *typical* weighting will probably be polyarchic.³⁴

(4) It seems that we might suggest with equal plausibility that those committeemen from whom everyone expects a great deal will be encouraged to work harder and, consequently, will be able to contribute more than anyone expects. But this reply merely reproduces the same unwarranted assumption on which the objection is based. That

is, there is no reason to assume that committeemen must be weighted long before the final vote on the issue is taken. Indeed, there is no doubt at all that the weighting should immediately precede the voting and that *both* should take place after a discussion of the issues. Moreover, every committeeman should be given an agenda far enough in advance of the meeting to provide some opportunity for him to become familiar with the issues. Given these tactics and the fact that our procedure allows committeemen to gain complete control of a committee decision, it would seem to be difficult for a person with average ability to justify an attitude of impotence.³⁵

(5) The application of a minimax loss strategy is not incompatible with the assignment of initial weights according to judgments of relative competence, i.e., it does not entail an anarchic weighting. If someone regards a less "correct" committee decision as a greater loss than a certain amount of personal influence on that decision, then the minimax strategy would lead him to distribute his initial weight in support of competence, rather than merely in support of himself. In effect then, he would be applying the strategy to the following alternatives:

- (a) Keep everything and obtain less "correct" committee decision.
- (b) Distribute part and obtain more "correct" committee decision.

The second alternative might well represent the smallest possible loss. The important unanswered question is: Will enough people value "correctness" more than personal influence to make our procedure effective?

(6) The evidence for these suggestions is fairly strong. So all we can do is try to devise ways to minimize the effects of such propensities.³⁶ One way to tackle the problem is to construct a set of minimum standards according to which initial weights must be distributed. For example, in the case of the selection of a textbook, we might give

³² Lester Coch and John R. P. French, Jr., "Overcoming Resistance to Change," *Human Relations*, vol. 1 (1948), p. 532.

³³ William R. Dill, "The Impact of Environment on Organizational Development," in Sidney Mailick and Edward H. van Ness (eds.), *Concepts and Issues in Administrative Behavior* (Englewood Cliffs, N.J., Prentice-Hall, Inc., 1952), p. 105.

³⁴ This is suggested, but by no means proved, by Dahl's research in New Haven, Conn., *op. cit.*, p. 228. A more extensive investigation by Robert E. Agger, Daniel Goldrich, and Bert E. Swanson suggests a number of other alternatives, *The Rulers and the Ruled* (New York, John Wiley and Sons, Inc., 1964), pp. 478-537.

³⁵ Robert E. Lane's investigation of the political attitudes of 15 "average" men led him to the conclusion that "people tend to care less about *equality* of opportunity than about the availability of *some* opportunity," *Political Ideology* (New York, The Free Press, 1962), p. 79.

³⁶ Guilford introduces techniques to estimate some of the types of errors we listed. They obviously require the assistance of a trained statistician. *Op. cit.*, pp. 280ff.

so many points for having taken so many courses, for having taught so many, for having published so many related papers, etc. Other more reliable tests would have to be developed for other types of issues. A second way to attack the problem would be to have the committeemen present their "credentials" with precise and standardized descriptions to reduce vagueness and ambiguity. This would reduce the temptation to interpret perceptions in a more or less arbitrary fashion.³⁷

(7) Insofar as events are bound to take place exactly as we prefer with or without our (formal) efforts, such efforts must be regarded as unnecessary. But the ultimate triumph of wisdom over ignorance is hardly inevitable. Hence, we seem to be obliged to take *some* steps to "further the cause of rationality and good decision-making." What I would like to suggest here is that the formalization of influence in some situations might be a step in the right direction.

(8) It seems clear that in some situations it would be wiser to *copy* a more competent member's voting behavior than to give him all or part of one's weight. However, in other situations, the latter strategy seems preferable to the former. For example, suppose that one is not only interested in the "correctness" of a committee decision, but also in appropriately allocating the rewards or punishments resulting from that decision. The distribution of weights could be used as a basis for fixing responsibility and for fairly allocating rewards and punishments. Similarly, weight distribution provides an explicit record of committeemen's views of one another's competence which could be useful for the choice of operating procedures and personnel. Again, it is often important to know whether a committee decision has received a certain amount of support on its own as it were or simply on the strength of its supporters. After all, to return to a point raised in Sect. IV, it is one thing to endorse a particular alternative and quite another to endorse a particular person who endorses that alternative. A ten-membered committee decision with one knowledgeable supporter and nine administratively competent but technically ignorant supporters seems to have less support in some sense³⁸ than the same decision with

nine knowledgeable supporters and one supporting administrator. While weight distribution (according to the procedure suggested here) might not alter the final number of points received for such a decision in either case, it would provide some intuitively useful extra bits of information whose value would tend to vary directly with the error and implementation costs of any decision. Of course, one could probably gather the extra information by additional research of some sort, but weight distribution always provides such information and always makes the subtle difference in the *kind* of support received by any alternative completely explicit.

VII. OBJECTIONS AND REPLIES: WEIGHTED VOTING

After reviewing the objections and replies to the procedure of weighting voters, it seems fair to say that as yet neither a case for nor against it has been decisively made. There are still a number of loose ends to be tied up, some of which require empirical investigation more than logical analysis. However, in this section we shall assume that our procedure for weighting voters is more or less acceptable and ask: How do the various schemes of weighted voting fare in the light of our five adequacy criteria?

Taking the least problematic point first, I think that we may grant the internal consistency [condition (3.2)] of all six schemes. It is certainly true that if one requires some sort of piecemeal comparison of alternatives (e.g., pairwise comparison, triplewise, etc.) and assumes that both individual and group systems of preferences must be transitive, then all of these schemes are liable to generate the so-called "paradox of voting" in certain situations.³⁹ But it seems to me that there is no good reason to expect group systems of preferences to be transitive or to require piecemeal comparisons of alternatives. So the "paradox" is something of a red herring.⁴⁰

It may be recalled that in Sect. V we showed that the total weight scheme is inefficient [violating (3.4)] in the sense that it is liable to be unable to take account of available information. Similarly, the weighted Borda-utility scheme must be

³⁷ There is no doubt at all that the (intentional and unintentional) ambiguity of politicians' claims is highly correlated with the erroneous interpretations of voters, e.g., see Berelson, Lazarsfeld, and McPhee, *loc. cit.*

³⁸ The precise nature and plausibility of this "sense" is still rather vague to me.

³⁹ An excellent review of the formal issues surrounding this paradox may be found in Yasusuke Murakami, *Logic and Social Choice* (London, Routledge and Kegan Paul, Ltd., 1968), chs. 5 and 6.

⁴⁰ I am in complete agreement with Dahl and Lindblom, *op. cit.*, pp. 422-424.

inefficient because Borda-utilities vary at regular intervals whether or not the preferences of committeemen vary that way. For example, if someone ranks D_1 above D_2 and the latter above D_3 , then the Borda-utility of D_1 is twice that of D_3 whether or not the committeemen believes D_1 is twice as "good" as D_3 . While we are not bound to use the whole numbers 0, 1, 2, 3, . . . , as Borda-utilities, we are bound to use some regularly increasing set (e.g., 0, $\frac{1}{3}$, $\frac{2}{3}$, . . . ; 0, $\frac{1}{4}$, $\frac{1}{2}$, $\frac{3}{4}$, . . .) because it is this very regularity that distinguishes this scheme from the total weighted-utility scheme. Hence, whenever committeemen have preferences that do not vary at regular intervals, they will be unable to "feed" this information into the weighted Borda-utility scheme. Hence, it too violates (3.4). Since the split weight Borda-utility scheme provides some means of breaking the rigid Borda-utility mold (namely, by appropriately splitting one's weight), it does not violate (3.4). Furthermore, for roughly the same reason, neither do the other three schemes.

The total weight scheme seems to be free of any peculiar problems of practicability [condition (3.3)]. However, as we saw in Sect. V, the total weighted utility, split weight utility, and split weight Borda-utility schemes are all impracticable because the labor costs of applying them are greater than the costs of applying the equivalent split weight option scheme. Furthermore, with the exception of the total weight scheme, they all presuppose (i) the ability of some committeemen to weakly order their preferences for various alternatives and (ii) the interpersonal comparison of utility. Both of these assumptions have been debated so exhaustively in the literature that it seems very unlikely that anything novel can be said here.⁴¹ So, I shall restrict my "defense" to the following brief remarks. In defense of (i) it should be noted that people are frequently (though by no means always) able to weakly order their preferences, and that this must be regarded as favorable evidence for the view that, as far as this problem is concerned, our voting schemes will frequently be practicable.⁴² In defense of (ii) it must be admitted that the very fact that fairly stable wage and price systems exist may be taken as evidence that some services and

commodities must have roughly the same value for many people. That is, if people of roughly equal means are willing to pay the same price for a certain commodity or do the same work for the same salary, then it seems more likely that they are receiving the same rather than different amounts of satisfaction from the exchange. Hence, this second problem seems to be primarily one of constructing appropriate technical "devices" to obtain reliable measures of comparability. Nevertheless, neither economists nor psychologists have been able to produce such "devices" after many years of trying. So it would be gratuitous to assume that these schemes did not face a serious problem of practicability.

The total weight scheme seems to violate condition (3.1), freedom from coercion, because it does not allow committeemen to distribute the weight of their votes in accordance with their preferences. They might then, be forced to give more support to certain issues than they prefer or simply withhold all of their influence. Similarly, the weighted Borda-utility scheme is liable to be coercive when it is suppressing information.⁴³

The split weight option scheme and its equivalents may be shown to violate (3.1) as follows. We have a three-membered committee in which each member has a weight of unity and a vote is to be taken to select one of four alternatives. If each member voted his "true" preferences, the result would be a victory for D_1 as follows:

	D_1	D_2	D_3	D_4	W_i
M_1	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{6}$	0	1
M_2	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{6}$	0	1
M_3	0	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	1
Total	1	$\frac{2}{3}$	$\frac{5}{6}$	$\frac{1}{2}$	

However, M_3 figures that "half a loaf is better than none" and that while his first choice is a lost cause, his second choice would win if he put all his weight behind it. By misrepresenting his preferences thus,

	D_1	D_2	D_3	D_4
M_3	0	0	1	0

he insures a victory for D_3 , because the new total is:

D_1	D_2	D_3	D_4
1	$\frac{1}{2}$	$1\frac{1}{2}$	0

⁴¹ I have entered the periphery of the arena in "Postulates of Rational Preference," *Philosophy of Science*, vol. 34 (1967), pp. 18-22, and "Estimated Utility and Corroboration," *British Journal for the Philosophy of Science*, vol. 16 (1966), pp. 327-331.

⁴² It is worthwhile to note, I think, that the chances of performing this task seem to be increased by considering Bentham's seven aspects of the "satisfaction" attached to a particular decision, viz., intensity, duration, certainty, propinquity, fecundity, purity, and extent. On this point Wayne A. R. Leys' discussion in *Ethics for Policy Decisions* (Englewood Cliffs, N.J., Prentice-Hall, Inc., 1952) is very helpful, pp. 13-32.

⁴³ A similar objection was raised by Condorcet against Borda's numbers in 1785 according to Isaac Todhunter, *A History of the Mathematical Theory of Probability* (New York, Chelsea Publishing Co., 1949), pp. 433-434.

Hence, for M_3 to insure a victory for his second choice given the split weight option scheme and the particular distribution of preferences of the other members of the committee, he must vote contrary to his preferences. So, the split weight option scheme and its equivalents violate (3.1); which means that all of our schemes are liable to be coercive.

Since the procedure for weighting voters is designed to guarantee the satisfaction of condition (3.5) and since it has rather precariously passed inspection, one might be inclined to say that the six schemes employing it are also provisionally acceptable on this score. Unfortunately, such an inference would not be warranted, because it is undoubtedly the case that all of these schemes are liable to violate (3.5). If, for example, every member of a three-membered committee has a weight of influence equal to unity going into an election and two of them abstain, the third member's weight of influence on the final outcome would jump from a third to one! But short of violating (3.1) there is no way to prevent such "anomalies" from arising. Hence, after all is said and done, we still do not have a scheme that will guarantee the satisfaction of (3.5). Indeed, we now know that it is impossible to construct such a scheme without violating (3.1). Thus, we have arrived at a conclusion that is analogous to Arrow's General Possibility Theorem.⁴⁴ That is, it is impossible to construct a voting scheme such that for *every* individual under *any* circumstances, both freedom from coercion and the distribution of influence according to relative competence are *guaranteed*. Of course, given certain individuals and circumstances, influence may well be distributed according to relative competence without coercion. So (3.1) and (3.5) are certainly consistent. But no procedure can *guarantee* their joint satisfaction.

If we let "*p*" be short for "passes" and "*f*" for "fails," then the views presented in this section may be summarized in the following matrix. The last

line has been added to facilitate comparison with what we have been calling the "usual (democratic) voting procedure."

Voting Scheme	Condition				
	3.1	3.2	3.3	3.4	3.5
total weight	<i>f</i>	<i>p</i>	<i>p</i>	<i>f</i>	<i>f</i>
split weight option*	<i>f</i>	<i>p</i>	<i>f</i>	<i>p</i>	<i>f</i>
total weighted-utility*	<i>f</i>	<i>p</i>	<i>f</i>	<i>p</i>	<i>f</i>
weighted Borda-utility	<i>f</i>	<i>p</i>	<i>f</i>	<i>f</i>	<i>f</i>
split weight utility*	<i>f</i>	<i>p</i>	<i>f</i>	<i>p</i>	<i>f</i>
split weight Borda-utility*	<i>f</i>	<i>p</i>	<i>f</i>	<i>p</i>	<i>f</i>
usual procedure	<i>f</i>	<i>p</i>	<i>p</i>	<i>f</i>	<i>f</i>

* indicates equivalent schemes

As the table discloses, the weighted Borda-utility scheme is the weakest of the lot, and all of the others are equally acceptable. At any rate, they are equally acceptable provided that one regards each of the five adequacy conditions as equally important. If the conditions were weighted unequally then one of the voting schemes might prove to be superior to the others.

VIII. CONCLUSION

Insofar as we have succeeded in providing a more or less systematic basis for logical and empirical investigations of group decision procedures which permit formal or explicit inequalities among voters and votes, our more general task has been completed. Insofar as we have been able to throw some light on the relationships, strengths, and weaknesses of various more or less plausible solutions to the problem of selecting an acceptable group-decision procedure, our more specific task has been accomplished. Perhaps the most appropriate summary of our results on the latter score would be a loose paraphrase of a remark made by Sir Winston Churchill: "The usual (democratic) procedure is a very bad form of government, but it is every bit as good as all the others."⁴⁵

University of Pittsburgh

Received April 22, 1969

⁴⁴ Arrow, *op. cit.*, pp. 46-60.

⁴⁵ I would like to express my gratitude to Sidney J. Herzig, George Yoos, William Hughes, and especially Thomas A. Schwartz for their help and encouragement.

II. MUST ONE PLAY THE MORAL LANGUAGE GAME?

ALAN GEWIRTH

THE title of this paper may be criticized on the ground (with which I sympathize) that morality is a highly serious matter and hence not a game. Let me point out, then, that the question I am asking may also be put in such ways as these: Must one use (and not merely mention) moral concepts? Must one make any moral judgments at all? Must one act morally, or accept moral obligations, or be moral? To play the moral language game is to use moral concepts in all sincerity and seriousness, and such use involves an acknowledgment by the user that moral concepts apply and ought to apply to his own conduct as well as to that of others. Hence, while it will be advantageous to retain the linguistic emphasis marked by the title, there is no conflict between referring to the "moral language game" and recognizing the seriousness of morality.

I. ANALYSIS OF THE QUESTION

Three negative views on morality must be distinguished. The *nihilist* holds (whether on extreme positivist or other grounds) that such words as "ought" and "right," at least in moral contexts, are not meaningful at all. The *amoralist*, while granting that such words are meaningful, holds that they have no application, at least in his own case; he denies that there is any valid reason for him to use moral language or to act in accordance with the requirements of any morality. The *moral sceptic*, while admitting that there may well be reasons, even conclusive ones, for using moral language in general and hence for having some morality or other, holds that there is no way to justify any one specific morality as against any

other. The amoralist is hence more radical than the moral sceptic, but less so than the nihilist.

While nihilism is an obviously indefensible position, the case is otherwise with the amoralist, and it is to him that my question refers. I am asking whether amoralism can be refuted, that is, whether reasons can be given which conclusively justify an individual's playing the moral language game and which hence show that it is irrational for him to refrain from making moral judgments or acting in accordance with any morality at all. Many contemporary philosophers have given a negative answer to this question. The following statements are typical:

There can be no complete non-personal, objective justification for acting morally rather than non-morally.¹

[A man] may refuse to make any moral judgment at all, even one of indifference. . . . Now it will be obvious that in [this] case there is nothing that we can do. . . . Such a person is not entering the arena of moral dispute, and therefore it is impossible to contest with him.²

A man who does not care what happens to other people has no reason . . . for adopting *any* moral rules.³

For a thoroughly amoral intelligence, nothing in principle can serve as a reason for *inducing* him to accept any moral responsibilities. Metaphysical elaborations, logical arguments, empirical generalizations and data, and, finally, all moral discourse with its lavish, complex and ingenious devices of persuasion are wholly inadequate. No reasons are possible.⁴

Put succinctly, using moral language commits me to a moral point of view, but nothing commits me to using moral language. . . . "Is it absurd, or self-contradictory, to refuse to consider any issues as what they call 'moral issues'?" . . . When the skeptic's doubts are raised in this form, I do not see how they can be quieted. . . . It follows that there is not

¹ Kai Nielsen, "Why Should I Be Moral?", *Methodos*, vol. 15 (1963), pp. 297-298.

² R. M. Hare, *Freedom and Reason* (Oxford, 1963), pp. 100-101.

³ C. H. Whiteley, "Universalisability," *Analysis*, vol. 27 (1966-67), p. 49 (italics in original).

⁴ A. I. Melden, "Why Be Moral?", *The Journal of Philosophy*, vol. 45 (1948), p. 455 (italics in original).

necessarily any way in which we can *reason* a man into thinking morally, or rather into a commitment to *act* morally.⁵

It will have been noted that these statements refer to an *individual's* using moral language and accepting moral obligations. This is also reflected in the "one" of the title. My question is not whether a society or group of men must have a morality, but whether an individual must have or accept any morality, that of his society or any other. Many philosophers, while upholding the Hobbesian answer to the social question, have confused the issue by assuming that this answer also fits the individual question. An individual may, however, recognize and even rejoice that his society has a morality while refusing to accept it or any other for himself. Such an amoralist need not be refuted by Hobbes's argument that it is irrational (in the sense of imprudent and appealing to the empirically improbable) to count on being able to deceive one's fellows. As has been pointed out by elitists from Callicles to the present (not sharing Hobbes's *de facto* egalitarianism), the man of superior strength and cunning may flout with impunity the rules he upholds for others. Nor can he be convicted of inconsistency on this account; for, so far as has hitherto been shown, he need not be saying or thinking that other men *ought* to abide by these rules but only, at most, that he *wants* them to do so because it suits his own purposes. There is no contradiction between *X's* wanting other men to do *y* because it suits his own purposes and his not wanting to do *y* himself because this likewise suits his own purposes.

To say that such a person is immoral is, of course, to beg the question, which is whether there is any rational ground for holding that he must himself

use such moral concepts with respect to himself. Another line of attack, however, is that the man who has no morality is mentally ill and hence must be put away or given medical treatment rather than be taken seriously as a protagonist of an arguable position. Now it is true that "psychopaths" are sometimes defined in terms of "amoral" behavior.⁶ An amoralist need not, however, be a psychopath: he need not be "impulsive" or unable to control his acts in the light of consequences, and in fact he may act both prudently and benevolently. He is similar to the psychopath in not feeling shame, remorse, anxiety, or guilt, but this is not because he is unable to feel these but rather because he regards them as inappropriate or unnecessary.

It may be objected that if the amoralist can decide whether or not to have guilt feelings, then he must be a superman. A similar idea, however, is found in Freud's view of psychoanalytic therapy as aiming to release men from the tyranny of the superego with its guilt feelings introjected from childhood experiences: the mature adult can control these feelings instead of being at their mercy. Indeed, the general point goes back at least to Aristotle's conception of the "self-indulgent" man as one who *chooses* to pursue excessive pleasures and feels no remorse at having done so; and similarly with the "shameless" man (*Nicomachean Ethics*, IV, 9; VII, 7,8). To be sure, there is a difference between occurrent and dispositional choices and controls, between controlling guilt feelings at the very time they might have occurred and the more long-range control over the factors that might lead one to have guilt feelings. Both kinds of control, however, may be attributed to the amoralist. He does not have guilt feelings because he holds that there are no sound reasons for having them.

It is this last feature that gives point to arguing

⁵ Anthony Ralls, "The Game of Life," *The Philosophical Quarterly*, vol. 16 (1966), pp. 27, 32 (italics in original). For similar statements, see H. D. Aiken, *Reason and Conduct* (New York, 1962), p. 86; A. P. Griffiths, "Justifying Moral Principles," *Proceedings of the Aristotelian Society*, vol. 58 (1957-58), pp. 104-105, 109; J. C. Thornton, "Can the Moral Point of View Be Justified?", *Australasian Journal of Philosophy*, vol. 42 (1964), p. 32; D. P. Gauthier, "Morality and Advantage," *The Philosophical Review*, vol. 76 (1967), p. 470; G. P. Henderson, "Moral Nihilism," in *Studies in Moral Philosophy, American Philosophical Quarterly Monograph No. 1* (1968), pp. 47ff.

⁶ See the "Psychiatric Glossary" published by the Committee on Public Information of the American Psychiatric Association, which defines a psychopath as: "A person whose behavior is preeminently amoral or anti-social and characterized by impulsive, irresponsible actions satisfying only immediate and narcissistic interests without concern for obvious and implicit social consequences, accompanied by minimal outward evidence of anxiety or guilt" (quoted in R. L. Jenkins, "The Psychopathic or Antisocial Personality," *Journal of Nervous and Mental Disease*, vol. 131 [1960], p. 318). See also Henry Cleckley, *The Mask of Sanity* (St. Louis, 1950, 2nd ed.), p. 545; Ian Gregory, *Psychiatry, Biological and Social* (Philadelphia, 1961), p. 452; F. J. Braceland and Michael Stock, *Modern Psychiatry* (New York, 1963), p. 334; A. A. Terruwe, *Psychopathic Personality and Neurosis* (New York, 1958), pp. 51-54. Also relevant here are the 19th century expressions used to define the psychopath as a "moral imbecile" and "moral defective," as well as current characterizations of him as showing "absence of ethical and moral appreciation," "loss of all ethical sense," and "affectional irresponsibility." See, e.g., E. A. Strecker, *Basic Psychiatry* (New York, 1952), p. 298; Gregory, *op. cit.*, p. 452; Norman Cameron and Ann Margaret, *Behavior Pathology* (Boston, 1951), p. 205.

with the amoralist, and it bears on the "must" of my question. The amoralist as I here conceive him (and as is also suggested by the references to "reason" in the last three quotations given above) professes to be guided by reason in that he is prepared to do that for which (logically) good justifying reasons can be given. He accepts the reasons of deductive and inductive logic, including the evidence of empirical facts. But he denies that these reasons justify or require that he make any moral judgments or accept moral obligations; and he also denies that there is any other "rationality" (including a distinctively moral one) which would rationally justify these things. My question, then, is whether, given the deductive and inductive reasons which the amoralist accepts, he also rationally must, in virtue of accepting these reasons, accept for himself the use of moral language and the corresponding moral obligations. This question is hence not open to the usual charge brought against questions of the "Why should I be moral?" type, that they are circular because their use of "should" already involves accepting moral reasons and hence being moral. The "must" of my question is a logical (deductive and inductive) justificatory "must," not a moral one (nor, of course, a physical one involving force or coercion).

Nor is my question open to the Humean charge that it asks for a (logically impossible) derivation of a "practical" (moral) commitment from the purely "theoretic" reasons of deductive and inductive logic. For the amoralist's commitment to these reasons is practical as well as theoretic: he is already disposed to accept and act upon that to which the weight of deductive and inductive reasons leads; and he is also conatively "normal" in that he has the self-interested motivations common to most men, and is willing to expend the effort needed to fulfill them. (It will be in this morally neutral sense that I use the phrase "rational and normal person" in this paper.) The question, then, is whether, given this rational and conative equipment alone, he must necessarily, despite even an obdurate initial rejection of morality, play the moral language game and accept moral obligations.

But what is meant by "moral" in expressions like "the moral language game"? It must be emphasized that I am using "moral" in the

general sense in which it is opposed to "non-moral," not to "immoral." In asking, then, whether the amoralist must play any moral language game at all, we must have some elucidation of the general concept of a morality, a notoriously difficult task. W. K. Frankena has provided a convenient summary of recent attempts at such elucidation. According to this, an individual *X*'s "action-guide" is a moral one or constitutes a morality if and only if it satisfies such criteria as the following:

- (A) *X* takes it as prescriptive.
- (B) *X* universalizes it.
- (C) *X* regards it as definitive, final, over-riding, or supremely authoritative.
- (D) It includes or consists of judgments (rules, principles, ideals, etc.) that pronounce actions and agents to be right, wrong, good, bad, etc., simply because of the effect they have on the feelings, interests, ideals, etc., of *other* persons. . . . Here "other" may mean "some other" or "all other."

This list may be criticized on various grounds. Some philosophers have emphasized one or another of these criteria as basic; and it is by no means certain that the five writers quoted above had in mind precisely these criteria of "moral." Since, however, my argument will be less open to charges of undue simplification if it recognizes more rather than fewer requirements as having to be satisfied, I shall adopt the whole list as jointly constituting the criteria of someone's having a morality and using moral language.

II. FROM "IS" TO NON-MORAL "OUGHT"

The heading of this section must be understood in a second-order rather than a first-order sense. I am not claiming that an "ought"-statement is to be derived from an "is"-statement, but rather that a proposition which says that someone makes an "ought"-statement (or has an "ought"-belief) is entailed by "is"-statements which set forth certain facts about his desires and his other beliefs. Specifically, I shall argue that, given the facts (1) that a rational and normal amoralist *X* wants to have something *y*, (2) that he believes that his doing *z* is a necessary and sufficient means to his having *y*, (3) that he believes that doing *z* involves

⁷ W. K. Frankena, "The Concept of Morality," *The Journal of Philosophy*, vol. 63 (1966), pp. 688, 689 (italics in original). Frankena has also discussed this subject in two other valuable papers: "Recent Conceptions of Morality" in H. N. Castaneda and George Nakhnikian (eds.), *Morality and the Language of Conduct* (Detroit, 1963), pp. 1-24, and "The Concept of Morality," *University of Colorado Studies, Series in Philosophy*, No. 3 (1967), pp. 1-22. While these two papers are more detailed on several points, they present no important differences of doctrine from that of *The Journal of Philosophy* paper.

a choice between alternatives and carries a "price" in the sense of some effort and constraint on his part, (4) that he believes that doing z is in his power, and (5) that he does not believe there is any superior counter-consideration to his having y or doing z , it logically follows (6) that he believes that he ought to do z . Note that I say that (6) "logically follows"; I am not merely making a contingent prediction about X 's psychology. (6) logically follows from (1)–(5) because the use of an "ought"-statement (or the having of an "ought"-belief) just is the way in which a rational and normal person shows his awareness of the constraints or requirements involved in what he believes he must do to attain something he wants. I am not saying that all "ought"-statements or "ought"-beliefs are of this instrumental sort, but that at least some are.

Let us designate by 'O' statements of the form "I ought to do z ." What is it for X to use or say O? Relevant here are two of the previously listed criteria for someone's having a "moral action-guide." First, "X takes it as prescriptive" [criterion (A)]; that is, if X uses O, then he gives an at least *prima facie* endorsement to his doing z and hence regards doing z as a requirement for or constraint on his conduct. Second, X believes that this requirement or constraint is legitimate or valid—in other words, he regards what O expresses as authoritative—in that he believes that there are good or justifying reasons for his doing z . (These need not be *morally* good or justifying reasons). This latter point is obviously related to criterion (C) above, save that criterion (C) referred to "supremely authoritative." I shall deal with this distinction subsequently.

It is because these features of prescriptiveness and authoritativeness are involved in X 's desires and beliefs (1)–(5) listed above that the latter entail his having the "ought"-belief or making the "ought"-statement listed at (6). For since X believes that his doing z is necessary to his having y , which he wants, and he does not believe that there is any superior counter-consideration, he to that extent endorses his doing z , regarding it as a requirement for his conduct justified by reason of his wanting y . For him to regard in this way his doing z is for him to believe that he ought to do z .

This thesis and its accompanying argument may be clarified if we consider some of the objections that may be brought against them.

(a) I seem to be saying that the statement " X believes he ought to do z " is exhaustively derivable from factual statements about X 's desires and his other beliefs. But this commits the naturalistic fallacy in one of its familiar forms. Moreover, since the concept of "ought" signifies what goes counter to our inclinations, how can it, or a belief about it, be derived from those very inclinations (or desires)?

(b) Since there is a difference between what one wants and what one ought to want, it is illegitimate to base an "ought"-conclusion on one's wants *per se* and on the means and constraints necessary for their satisfaction. Both the wants themselves and the means to their satisfaction must first be scrutinized to see that they do not conflict with any more pressing obligations.⁸

(c) It is therefore incorrect to say that the factual statements (1)–(5) listed at the beginning of this section entail the statement (6) that " X believes that he ought to do z ." What they entail is much more modest: either (i) simply that X 's doing z is a necessary or sufficient condition of his having y , or (ii) that X wants to do z as a means to his having y , or (iii) that X accepts some such singular imperative as "Let me do z ," or (iv) that X says, "So I'll do z ," or (v) that X says, "If I'm smart I'll do z ."

My answer to (a) is that while the concept of "ought" cannot be defined in terms of desires alone, its authoritative and prescriptive features, so far as they enter into a person's "ought"-beliefs, are derivable from the combination of the person's desiring something, his believing that that desire provides a reason or justification for action, and his being aware of the requirements or constraints which the satisfaction of that desire imposes on his other desires and acts. Since the derivation is in part from beliefs about "normative" factors like justificatory reasons and requirements, there is a sense in which it is not naturalistic. In any case, it is false to say that beliefs about the satisfaction of desires or inclinations cannot refer to restraints on or control of other desires. The belief that one ought to go to the dentist is an obvious counter-example.

In answer to (b), it must be recalled that what I

⁸ See the criticisms directed against Max Black, "The Gap Between 'Is' and 'Should,'" *The Philosophical Review*, vol. 73 (1964), pp. 125–181, by M. F. Cohen, "Is and 'Should': An Unbridged Gap," *ibid.*, vol. 74 (1965), p. 224, and T. Y. Henderson, "The Gap Between Good Strategy and Right Action," *Philosophy*, vol. 41 (1966), pp. 260–267. As is indicated both at the beginning of this section and in my reply to (b), below, my project does not, like Black's or Searle's (see next note), involve deriving an "ought"-statement from an "is"-statement.

am claiming to derive from factual premisses about *X*'s wants is not an "ought"-conclusion *tout court* but rather a conclusion that *X* must have a certain "ought"-belief, given his wants and certain other beliefs of his. It is one thing to say that he must have this "ought"-belief; it is another thing to say that the "ought"-belief is itself justified, all things considered. For the latter statement to be correct, it would be necessary to consider the nature of *X*'s wants and their relation to the wants of other persons. This, however, is not part of my present concern, nor is it necessary to my present argument. It must also be recalled that my derivation of *X*'s "ought"-belief included the statement that "he does not believe there is any superior counter-consideration to his having *y* or doing *z*."⁹ I used the vague expression "superior counter-consideration" to refer to any factor that might weigh more heavily with *X*, including any of his other desires or beliefs. Such weighting of factors is, of course, a basic feature of calculative reasoning. Now so far as concerns an amoralist's explicit awareness, these counter-considerations do not include any moral obligations. But it would be incorrect to conclude from this that it is logically illegitimate for him to use the concept of "ought" at all. This concept is neutral as between conflicting practical frameworks; as I shall argue below, it has a common core of meaning both for deontologists and non-deontologists, as well as within moral and non-moral value systems. In its first-person use "ought" means that the user acknowledges a requirement for or constraint on his conduct, justified by reasons.

This meaning, entailed by statements (1)–(5) listed above, is not conveyed by the proposed substitutes for the "ought"-belief listed under objection (c). Hence those substitutes fail to do justice to the full scope of the statements. (i) The statements present not merely a means-end relationship but reasons for pursuing the means, given that one wants the end. (ii) Nor do the statements imply merely that *X* wants the means in order to attain the end, as against recognizing that he ought to pursue the means. He may not want the means at all; and even if he does, this is not the same as recognizing the requiredness and constraint

involved in there being reasons that justify the sacrifice of his immediate desires. This point also applies to (iii) the singular imperative and (iv) the declaration of a decision. As to (v), when *X* says, "If I'm smart I'll do *z*," this cannot be interpreted as a mere conditional prediction such as might be uttered by someone who "plays it cool," as if it concomitantly also implied: "If I'm not smart I won't do *z*." Such indifference fails to catch the full force of someone's genuinely wanting something and recognizing and accepting the constraints needed to get it. Rather, the phrase, "If I'm smart," must be interpreted to mean: "If I have the intelligence I ought to have..." Hence, "smart" reintroduces the normativeness or requiredness for which the intended alternative was to be a substitute.

What I have tried to establish thus far, then, is that even an amoralist, so long as he is rational and normal in the senses indicated above, must use the concept of "ought." A person who did not use this concept would not be aware that any requirements or constraints were ever set for his conduct for any reason whatever, including his own self-interested desires. Even if the latter were his only desires, he would not be able to distinguish what he ought and ought not to do with a view to satisfying them. So soon as any person begins to deliberate between alternative courses of action to achieve any purpose of his, and thereby rules out some alternatives and accepts others, he necessarily uses the concept of "ought."

III. FROM NON-MORAL "OUGHT" TO MORAL "OUGHT"

According to Prichard, there is "a total difference of meaning" between the moral and the non-moral "ought," because the latter is exhaustively translatable into the idea of what is necessary to realize an agent's purpose whereas the moral "ought" has no reference to the agent's purpose.¹⁰ For the difference in meaning to be "total," however, a necessary condition would seem to be that the non-moral and moral "oughts" have no meaning in common save perhaps at an extremely general level, just as, for example, "race" meaning

⁹ This statement is similar to but not identical with the *ceteris paribus* clause in J. R. Searle, "How to Derive 'Ought' from 'Is'," *The Philosophical Review*, vol. 73 (1964), pp. 46ff. As his critics pointed out, for this clause to serve its purpose in his attempt to derive the obligation of promise-keeping from factual premisses, the clause must itself be evaluative (James and Judith Thomson, "How Not to Derive 'Ought' from 'Is'," *ibid.*, vol. 73 [1964], pp. 512ff.). This criticism does not apply to my above argument, however, because the "superior counter-considerations" are only those which the amoralist already accepts, so that they do not represent a covert "moral" assumption opposed to his own premisses.

¹⁰ H. A. Prichard, *Moral Obligation* (Oxford, 1949), pp. 90–91.

a biological classification and "race" meaning an athletic contest have nothing in common save as somehow pertaining (although in different ways) to living phenomena. But Prichard does not think that the moral and non-moral "oughts" are so extremely distant in meaning. For he admits that both "oughts" are imperatives. Now within the general realm of language, "imperative" signifies a much more specific characteristic or function than the two "races" signify in common. If we add to imperativeness (or prescriptiveness) the feature of authoritativeness, it seems more plausible to hold that the moral and the non-moral "oughts" share a central core of meaning, and that the moral "ought" adds to this the other features listed above: universalization, social concern, and supreme authoritativeness. Our task, then, is to show that at least some of the amoralist's uses of "ought" must exhibit these additional features.

The phrase "*X* universalizes his 'ought'-statements" may be misleading if it suggests that the universalizability of an "ought"-statement is at the option of the speaker. For a singular "ought"-statement necessarily implies a universal statement, regardless of whether the speaker admits this; and insofar as he is rational he necessarily admits it: the implication is so direct that it requires no extensive calculation. Specifically, *X*'s statement, "I ought to do *z* because I want to have *y*," is enthymematic; it entails the suppressed major: "All persons who want to have *y* ought to do *z*." The basis of this entailment is that, as we have seen, *X*'s "ought"-statement rests on the reason given in the "because"-clause; and it is a logical feature of all reasons that they are implicitly general, referring to a general rule or principle that serves to ground the connection asserted in the particular case.¹¹ Such a connection must hence obtain in all other cases to which the same rule or reason applies. In *X*'s full statement as given above, he is saying that his possessing the predicate *O* ("ought to do *z*") is justified by, has its reason in, his possessing the predicate *H* ("want to have *y*"). *X* is therefore logically committed to accepting the universal proposition that whoever possesses *H* possesses *O*.

Problems arise about the scope of this univer-

salization. Since the speaker himself provides the reason justifying his "ought"-statement, what is to prevent him from so individualizing that reason as to restrict its application to himself alone? When *X* says, "I ought to do *z* because I want to have *y*," he may ward off its implication that "all persons who want to have *y* ought to do *z*" by adding further qualifications in the "because"-clause: for example, "because I am over six feet tall, born January 1, 1945, on Harper Avenue, Chicago, and named *X*." The only universalization to which *X* would be logically committed would hence be: "All persons who want to have *y* and who were born on Harper Avenue and are named *X* . . . ought to do *z*."

The natural reaction to this piling up of "reasons" is: "What on earth does your being born on Harper Avenue or your being named *X* have to do with whether you ought to do *z*?" (I shall call this question '*Q*'). The implication of *Q* is that a justifying reason for an act must not include conditions which are irrelevant in the sense of unnecessary to the connection between the reason and the act for which it is a reason. *X* might reply, however, that his only reason for the act is that it is he, *X*, who wants to have *y*; and there is a necessary connection between (i) the act: someone named *X* does *z*; and (ii) the reason: that same person named *X* has or gets *y*, which he wants. Nevertheless, the point of *Q* still remains if we reformulate it as follows (*Q*₁): "What if your name weren't *X*? Wouldn't it still be the case that you ought to do *z* in order to have *y*?" As *Q*₁ indicates, *X*'s having the name he has is irrelevant to the necessary connection between (i) and (ii), for the connection would still obtain even if the words "named *X*" were omitted. To put it formally: when reasons *R* justify an action *A* by stating that *A* is a necessary condition for achieving some end *E*, *R* may not include any facts which are such that, on their elimination, the necessary connection between *A* and *E* would remain unchanged.

It may still be objected, however, that *X*'s reason for doing the act is so irreducibly egocentric that it cannot be expressed without explicit personal reference to him; hence it cannot be generalized

¹¹ It may be objected that in fields like history and law the giving of reasons for actions, or the use of "because" in causal explanations, may make no reference to any generalizations. (See William Dray, *Laws and Explanation in History* [Oxford, 1957] and H. L. A. Hart and A. M. Honoré, *Causation in the Law* [Oxford, 1959].) It is important, however, not to confuse the logical question of whether a generalization is implied by a "because"-statement with such pragmatic questions as whether historians or lawyers always use these implications in their work or whether they are able fully to ascertain which one of many different implied generalizations states the sufficient condition of a specific event. Negative answers to the pragmatic questions do not necessitate a negative answer to the logical question. Implicit admissions of this distinction can be found in Hart and Honoré, *ibid.*, pp. 14-15, 20-21.

so as to be the "same reason" as any other persons may have for performing their respective acts. "My reason for doing z is simply that I want that I have y , and this reason is different from any other persons' wanting that they have y . I don't care about anyone else's wants but only about my own; and nobody else has my wants." What is here claimed is that there is no way, by logic alone, to eliminate indexical expressions like "my" or "I" from the reason X has given for his doing z . If X 's reason can be universalized at all, it must be in a way that retains the individualizing reference to his own personal wants: "All persons who have my want that I have y ought to do z ."

The solution of this difficulty requires a closer look at the concept of wanting. Three elements must be distinguished: the person who wants something, his wanting, and what he wants. "Want" as a noun may refer to either of the last two. Now just as one person is numerically different from another, so are their respective wantings; in this sense the "wants" of one person are never numerically the same as those of another. They may, however, be generically or qualitatively the same insofar as the persons may want something in the same (or similar) way, as indicated, for example, by the kinds of effort they put forth to obtain whatever objects they want. If we look next at their wants in the sense of what they want, these too may be the same. Here, however, two alternatives must be distinguished. When we talk of what someone wants, this "what" may be expressed either as the object itself which is wanted or as the combination of the object and the person who wants it. In the first way, we say " X wants (or wants to have) y "; in the second way, " X wants that he have y ." Now in the first way it is clear that X and someone else (call him " W ") may want the same thing, so that their "wants" are the same. This point holds regardless of such further complexities as that X and W may want either the numerically same lounge chair or a lounge chair; in the latter case they still want the generically same thing, so that "wanting to have a lounge chair" serves as the common or same reason for the acts they perform to obtain one. But in the second way, when it is said that X wants that X have y and W wants that W have y , it may be thought obvious that what X and W want are not the same. And

indeed they are not numerically the same. But they may still be generically or proportionally the same, in that X 's having y is to X as W 's having y is to W . In each case y is the object of X 's and W 's similar desiderative attitudes, which serve as the same or common reason for their respective efforts to obtain y . To put it in other terms: X 's statement, "I want that I have y ," is a token of a type which can be truly and relevantly uttered by W and other persons as well. While each of these tokens has a different reference so far as concerns the 'I', they all have a common meaning in that they express the qualitatively same desiderative attitude toward the (numerically or generically) same object. Hence, those other persons' reasons for doing z are the same as X 's reason. To deny this would be like holding that one man's reason for doing something can never be the same as another man's because the thoughts in the heads of the two men are numerically different. But such a difference is irrelevant to the concept of the "same reason."

I have now considered and rejected two kinds of individualizing restrictions (deriving from proper names and first-person indexical expressions) on the universalization of X 's reason for believing that he ought to do z . More generally, as I have argued elsewhere,¹² even if X 's reason signifies a unique property which belongs only to him, still that property is similar or proportional to the properties of other persons, so that X 's reason for doing z must apply in a similar or proportional way to those other persons. I shall henceforth use ' S ' and ' U ' to refer, respectively, to singular and universal "ought"-statements of the form "I ought to do z because I want to have y " and "All persons who want to have y ought to do z ," where the latter are derived by universalizing the former in the way discussed above.

Thus far, I have tried to show that X necessarily accepts U , because as a rational and normal person he necessarily makes judgments of form S and necessarily accepts the universalizations which S entails. The claim I now want to make is that statements of form U are moral judgments. To establish this, I must show that U fulfills the four criteria listed above. We have already seen that it fulfills criterion (B), in that X derives it by universalizing S . (I assume that the universalization requirement may be fulfilled either by a singular

¹² I have discussed the "individualizability objection" to the universalizability thesis more fully elsewhere. See A. Gewirth, "The Non-Trivializability of Universalizability," *Australasian Journal of Philosophy*, vol. 47 (1969), pp. 123ff. I have also discussed the "criterion of relevant similarities" in "The Generalization Principle," *The Philosophical Review*, vol. 73 (1964), pp. 237ff.

judgment which one universalizes or by the resulting universal judgment.) Let us now turn to criterion (D), according to which a judgment to be a moral one must "pronounce actions . . . to be right [or] wrong . . . simply because of the effect they have on the feelings, interests, ideals, etc., of other persons." In view of the importance of this criterion for our question, we must try to elucidate it properly.

In the first place, criterion (D) refers to moral judgments as pronouncing actions to be "right [or] wrong," whereas my above argument has been in terms of "ought." This, however, raises no difficulty, for in any practical or prescriptive context where "ought" is used, "right" or "wrong" may also, *a fortiori*, be used: "ought" and "right" are related as subalternant and subalternate, so that "X ought to do z" entails "it is right that X do z," although not conversely. This entailment would be broken only if specifications of different contexts or reasons were introduced for the two concepts: if, for example, "ought" were used in connection with prudential or technical reasons and "right" in connection with legal reasons. Such differences of context, however, do not figure in my argument.

Secondly, when criterion (D) says that a moral judgment must "pronounce actions . . . to be right [or] wrong . . . simply because of the effect they have on the feelings, interests, ideals, etc., of other persons," the "effect" in question is obviously intended in such a way that an action is pronounced right if it promotes other persons' interests and wrong if it frustrates them, rather than the reverse. The aim of criterion (D) is to emphasize the socially-considerate or social-beneficial requirement of morality, just as when philosophers have raised the question, "Why should one be moral?" they have usually meant, "Why should one support, or act so as to promote, the interests of persons other than oneself?"

Third, although criterion (D) refers directly not to actions but to "judgments" which "pronounce" on actions, the criterion also requires for its fulfillment that one act in appropriate ways in accordance with one's judgments. This is especially clear from the fact that criterion (D) is sequential on criterion (A), according to which "X takes [his action-guide] as prescriptive." The familiar problems of *akrasia* which arise in connection with prescriptivist theories of the relation between moral concepts and action need no special attention here. As was indicated above in my discussion of "ought," when X says, "All persons who want to

have y ought to do z," this involves, insofar as his "ought" is prescriptive, that he endorses other persons' doing z, he urges them to do z, he commits himself to act in support of his endorsement, and so forth.

With these elucidations out of the way, I now wish to argue that U fulfills criterion (D). It will be recalled that statements of form U were necessarily accepted by X because, as a rational person, he recognized that they were entailed by his singular statements of form S ("I ought to do z because I want to have y"). Now in accepting statements of form U, "All persons who want to have y ought to do z," X holds that these persons' doing z is right because it will enable them to further or satisfy their respective interests in having y. To be sure, he is not saying that this is right "simply" for this reason, for his accepting the subordinate clause in U is consequent upon his accepting the "because"-clause in S. Since, however, the derivation of U from S is logically necessary, X is ineluctably committed to endorsing other persons' pursuit of their respective interests on the same basis as he endorses his pursuit of his own; he commits himself to accepting on the part of others the same kinds of interest-furthering acts that he upholds for himself. This generalized acceptance also means that X must look at his own proposed act in the light of the possible general performance of similar acts, since by U he declares such general performance to be justified.

Against my claim that U fulfills criterion (D) and is to that extent a moral judgment, it may be objected that there are important differences between the meaning of "ought" in U and the moral meaning of either "ought" or "right." Just as someone who is concerned with other persons' interests only because this causally advances his own interests is held to make prudential rather than moral judgments, so someone whose "ought"-statements (as in U) are concerned with other persons' interests only as a logical consequence of his own self-interested "ought"-statements may be held to be making not a moral judgment but a "derivatively-prudential" one. How can logic alone accomplish the transition from a self-interested "ought" (as in S) to a socially-concerned "ought"? There is, moreover, a difference between the way ends are viewed in U and in a moral judgment [in the sense of criterion (D)]. If someone makes a moral judgment that certain acts are right because they advance the interests of persons other than himself, he implicitly endorses the interests or

ends themselves: the acts or means are right because the advancement of those interests is right. But when X says (in U) that all persons who want to have y ought to do z , he is not endorsing their having y or holding that it is right that they have y ; he is not committing himself about their ends at all, but is saying only that if they want to have y , the right way to get it is by doing z , where "right" means merely technically efficient. Hence, U is only a technical judgment, not a moral one in the sense of criterion (D).

In answer to these objections, it is important to keep in mind that the question of U 's fulfillment of criterion (D) involves what X must accept as justified on the basis of the logical canons to which we have assumed him to be committed. The "transition" accomplished by these canons is simply from an "ought"-predication justified by a certain reason in X 's own case to its being justified in all the other cases to which that same reason applies. Such a transition from the individual to the social or universal is no more mysterious in the sphere of practical judgment than it is in the inference from a "mortality"-predication justified in the case of Socrates by reason of his being human to the same predication's being justified in the case of all other humans. It must also be remembered that the meaning of U (including its use of "ought") is to be understood not in isolation but in the light of U 's logical derivation from S . This point must be emphasized here and throughout our consideration of U : U is not merely a means-end statement toward which X may be neutral or indifferent; it is rather a personal-prescriptive statement by X . For since U is entailed by S , "ought" must have the same meaning in both statements; hence, since the "ought" in S is prescriptive, so too is the "ought" in U [thus fulfilling criterion (A)]. Moreover, in saying that all persons who want to have y ought to do z , X is endorsing their doing z for the same reason that he endorses (in S) his own doing z . Now since in S he accepts his wanting to have y as a sufficient reason or justification for his doing z , in U he must likewise accept other persons' wanting to have y as sufficient justification for their doing z . The point remains essentially the same if X bases his singular "ought"-statement on further reasons than simply his wanting to have y . In this case he would still regard his wanting to have y , combined with the other reasons, as justifying his doing z ; hence he would be logically required to regard other persons' wanting to have y , combined with

the other reasons, as similarly justifying their doing z . In either case, X is logically required to regard other persons' wants not with indifference but as legitimating reasons or justifications for their respective actions; he endorses not only the means but also the ends.

A further question may be asked about U 's relation to criterion (D): do any interests whatsoever count as fulfilling the criterion? Consider, at one extreme, a statement like: "All persons who want their cars to run smoothly ought to change the oil every thousand miles" (U_1); and at the other extreme: "All persons who want to kill other persons without being detected ought to feed them arsenic" (U_2). Neither of these, it may be held, is a moral judgment even if derived from the corresponding singular judgment: not U_1 , because the interests of the persons referred to are not sufficiently far-reaching or important; not U_2 , because the would-be poisoners consider only their own interests and threaten grave harm to other persons. The objection to U_1 , however, really bears on criterion (D). We could say either that *any* interests of persons will count as fulfilling criterion (D) when someone else says that acts are right because they advance those interests; or that even if only "important" interests count, still each individual is the judge of what is important to him; or, finally, that criterion (D) is to be amended so that only what are generally agreed to be important interests are to count, such as those bearing on physical or mental well-being. On either of the first two alternatives U_1 would satisfy criterion (D); on the third alternative it might not, but then it would be easy to find other examples of S where X believes he ought to do something because it is necessary to his basic health or happiness, and these when universalized would satisfy criterion (D) as amended.

As for the objection against U_2 , it must be remembered that the salient consideration in determining whether U_2 fulfills criterion (D) is not whether the poisoners' interests are harmful to others but rather whether X in making or accepting the judgment is holding that certain acts are right because they further the interests of persons other than himself—in this case, the interests of the poisoners. That these interests, and hence also X 's judgment endorsing them, are immoral may itself be a sign that the judgment is a moral and not a non-moral one; for one might argue that only what is moral (as opposed to non-moral) can be immoral. But even apart from such an argument,

it must be recognized that there are many moralities other than universalist or equalitarian ones; as Frankena points out, "criterion (D) allows for nationalistic and class moralities, for Nazism, for unequalitarianism."¹³ I take it that the position upheld by U_2 would fit among such moralities. In general, all the U -statements which X is logically committed to accepting involve the mutuality of his endorsing other men's pursuits of their respective interests on the same basis as he endorses his pursuit of his own.

Another consideration showing that U , regardless of its specific content, fulfills criterion (D) can be found in the connection of both U and S with the idea of mutual freedom. X 's statement "I ought to do z " entails "I ought to be free to do z ," where "to be free" means not to be interfered with by other persons in doing z . For if it is right that other persons interfere with X 's doing z or prevent him from doing z , then it is false that he ought to do z . Similarly, "all persons who want to have y ought to do z " entails "all persons who want to have y ought to be free to do z ." But anyone who accepts such universal statements as personal-prescriptive ones endorsing the end as well as the means, as we have seen that X must accept them, must also accept the singular statement, "I ought to refrain from interfering with other persons' doing z if they want to have y ." Hence, on the basis of his own self-interested acts and judgments, X must uphold a system of mutual rights and duties in which his "ought"-judgments in support of his own freedom of action entail further "ought"-judgments in which he accords similar freedom to others.

In the argument just given, it must be emphasized that no moral claim on other persons is directly involved in X 's statement, "I ought to be free to do z ," or in its equivalent statement "other persons ought not to interfere with my doing z ." These "oughts," like that in "I ought to do z ," derive only from X 's egocentric wanting to have y ; and this is also true of the "right" in "it is not right that other persons interfere with my doing z ." The moral claim [using "moral" in the sense of criterion (D)] comes in subsequently when X universalizes his singular judgment so that it becomes: "All persons who want to have y ought to be free to do z ." This universal judgment fulfills criterion (D) in at least two respects. First, it declares other men's unimpeded pursuit of their interests to be right because of its positive effect on their respective interests. Secondly, it commits X

himself to not interfering with or preventing these other men's doing z , for if X accepts that other men ought to be free to do z , then, by the definition of "free," he admits that he ought not to prevent their doing z . But this moral claim in the universal case involves no ambiguity in the case of X 's singular judgment "I ought to be free to do z ," for this latter is focused only on X 's own interests; hence, it does not itself make a moral claim.

It may be objected that "I ought to do z " does not entail "I ought to be free to do z ." For "do z " is ambiguous as between attempt and achievement. In competitive situations where both X and W want to have y and each believes he ought to do z in order to get y , it might well be the case that they cannot both do z (for example, marry the boss's daughter). Hence, since "ought" presupposes "can," X would believe only that he ought to try to do z . But this belief does not entail that X ought to be free to succeed in doing z so that W or other persons ought not to interfere with X 's succeeding in doing z . What it entails is rather that X ought to be free from interference in trying to do z .

Even if we grant this objection, it still involves an acceptance by X of the need for mutual freedom of action in the sense of attempt. That is, X must accept the judgment that other persons ought to be free to pursue (even if not to achieve or attain) their interests or objectives, on the same ground as he endorses this freedom for himself. And this judgment fulfills criterion (D).

In addition to the negative responsibility of non-interference, it can be shown in a parallel way that X 's initial "ought"-statement also logically commits him to accepting the positive responsibility of helping other persons. For suppose X 's doing z is impossible without other men's providing various kinds of essential conditions or services, which I shall call p . Hence, when X says "I ought to do z ," he must also accept the statement, "Other men ought to do p ." For since "ought" presupposes "can," if it is right that other men not do p , without which X cannot do z , then it is false that X ought to do z . To put it otherwise, if one endorses some end, then one must also endorse the necessary means to that end, at least *prima facie* or in the absence of superior counter-considerations. In addition, as we have seen, X must accept the general statement, "All persons who want to have y ought to do z ." And insofar as such men's doing z is impossible without other men's doing p , X must also accept that the latter men ought to do p . Now

¹³ "The Concept of Morality," *The Journal of Philosophy*, vol. 63 (1966), p. 692.

if *X* is included in this latter group, then he must concede that he too ought to do *p*. But obviously *X* is included in this group at least insofar as the "essential conditions" in question involve the maintenance in certain interpersonal relations of a certain modicum of honesty and trustworthiness. Hence, *X* must do his part toward maintaining such helpful conditions. This consideration also suggests that there are limitations as to the kinds of acts and wants that *X* may himself justifiably pursue, at least toward some persons.

The fact that *X* may be able to shirk these responsibilities, and in general to flout rules which he upholds for others, is here irrelevant. For once his initial statement that he ought to do *z* is given on the basis of his wanting to have *y*, the basis of the subsequent "ought"-statements to which he is committed is not prudential but logical. The reason why *X* must endorse other men's pursuit of their interests and accord them the needed freedom and aid is not the prudential or contingent one that if he interferes with or fails to aid other men's acts then he may probably expect them to interfere with or fail to aid his, but rather the logically necessary one that if a reason is held to justify an act because it is of a certain kind, then that reason must, on pain of contradiction, be held to justify all acts of that kind. There is, of course, much room for contingency in moral reasoning. What is at stake, however, in my insistence on the logical rather than prudential basis of *X*'s subsequent "ought"-statements is that if their basis were only prudential, then their validity or requiredness would merely be contingent on the degree to which *X* could not satisfy his interests or desires without endorsing and helping other men's pursuits of their interests. Hence there would be many kinds of circumstances in which *X* would simply have no moral obligations of the kinds described. Since, however, the basis of these obligations is logical, their validity is necessary independently of such contingent circumstances.

It must also be noted, however, that the rules which result from such obligations need not be, as with Hobbes, completely egalitarian and universalist: they still might, for example, provide for the hegemony of the physically or mentally stronger over the weaker. But the rules would at least set up a social-moral framework in which *X* would recognize that, on pain of self-contradiction, his acts in pursuit of his objectives must be affirmatively related to certain common standards bearing on the acts of other persons as well as his own.

Let us, finally, consider whether *U* fulfills criterion (C) (that *X* regards his action-guide as "definitive, final, over-riding, or supremely authoritative"). I said above that *X* regards his "ought"-statements as authoritative in that he holds that the requirements they impose on his conduct are justified by reasons deriving from his wants. But I also said that one of the factors entering into his statement or belief that he ought to do *z* because he wants to have *y* is that "he does not believe there is any superior counter-consideration to his having *y* or doing *z*." For if he did believe there was such a superior counter-consideration, he would regard it as removing the requiredness of his doing *z*. Does this mean that all "ought"-beliefs are regarded by the persons having them as supremely authoritative? Only in a relative sense, that is, as pertaining to the particular choice-situation in question. The phrase "supremely authoritative," however, is generic: it refers to a general range or kind of reasons as outweighing other kinds of reasons in respect to validity or legitimacy. Now it is quite conceivable that *X* might not even implicitly uphold such a hierarchy of reasons: he might regard first one kind of end or desire as most important, then a quite different and even conflicting one, and so forth. Clearly, however, such a policy (or rather lack of policy) would be irrational, for it could lead to frustration of the desires which, as a conative being, he is trying to satisfy. To avoid this, he must arrange his ends in an order of at least relative priority; which is to say that he must regard some of the reasons for his actions as supremely authoritative at least within a certain broad range of calculation.

It may be objected, however, that even if *X* regards some of his action-guides as supremely authoritative, he does not regard *U* in this way: hence his action-guides are not moral ones because those (like *U*) that fulfill criterion (D) do not fulfill criterion (C), and conversely. That *X* does not regard *U* as supremely authoritative emerges, first, from the point traditionally made by deontologists against teleological doctrines: since the "oughts" both in *S* and in *U* derive their authoritativeness for *X* from indicating the means to the ends that he wants to achieve, it must be these ends rather than the "oughts" that he regards as supremely authoritative. Since, moreover, *X* accepts *U* only insofar as it is a logical consequence of *S*, he must regard *S* as superior to *U* in authoritativeness.

The answer to these objections is that they prove too much. For if what is derivative can in no way be

regarded as supremely authoritative, if a morality as supremely authoritative consists only in the moral principle but not in the moral system which is justified by the principle, then no judgments, whether singular or general, could ever be moral ones so long as they were justified by a superior principle or reason. It seems more plausible, therefore, to hold that a morality consists not only in some general range of reasons contained in one or a few supremely authoritative principles but also in the system which is justified by those reasons or principles. It must also be kept in mind that authoritativeness is basically a matter of reasons or justifications, not only of motivations. Hence, the fact that *X*'s practical motivations, at least initially, may be exclusively self-interested rather than social does not preclude social concerns from figuring in what he regards as supremely authoritative.

I conclude, then, that *U* fulfills criterion (C) as well as the other three criteria of a moral judgment. The above arguments, if sound, have thus shown that any person, so long as he accepts the reasons of deductive and inductive logic and is conatively moral, must play the moral language game and have a morality. This result means, among other things, not only that amoralism is irrational (in a sense of "irrational" accepted by the amoralist himself as defined above) but also that to prove this there is no need to appeal to any peculiarly moral rationality, if there be any such.

Although I have interpreted the four criteria of "moral" listed above in such a way as to avoid any specifically "emotional" features, including conscience and sympathy (in criteria (C) and (D), respectively), *X* may well come to have such feelings as contingent consequences of his sincerely holding

practical beliefs which fulfill the criteria. Despite the great importance I myself attach to conscience and sympathy, I have not explicitly included them as essential to every morality, for a man may acknowledge obligations which he accepts as prescriptive, generalized, supremely authoritative, and concerned with other persons' interests, out of what he believes to be considered convictions about what is justified or right, even though he feels no sympathy for the other persons and regards conscience as at best unreliable and in any case as dispensable. It must also be remembered that what I have tried to show here is that *X*, the quondam amoralist, must have a morality, must play *some* moral language game. A further argument, using additional materials, is needed to differentiate, among the various possible moralities, the one which can itself be shown to be morally justified. Elsewhere¹⁴ I have tried to sketch such a further argument.

In conclusion, the whole project of this paper may be challenged. Why should we bother with the amoralist at all? Why shouldn't we be content to take the social point of view according to which a morality is required for the very existence of a society; why, in addition, should we take seriously the point of view of some deviant individual who would reject all moral language and hence his moral obligations to any society? I have two answers to such questions. One is that the deviant views of individuals deserve a hearing on their own merits. The other is that it is important to have shown that morality rests on a rational structure which can be used in a non-circular way to refute any attempt at a reasoned rejection of morality.¹⁵

University of Chicago

Received December 17, 1968

¹⁴ See A. Gewirth, "Categorical Consistency in Ethics," *The Philosophical Quarterly*, vol. 17 (1967), pp. 289-299.

¹⁵ For their comments on an earlier version of this paper I am indebted to various of my colleagues and students, and especially to Dr. Richard Parker.

III. AN EXAMINATION OF THE "SOUL-MAKING" THEODICY

CLEMENT DORE

I

THE so-called "soul-making" theodicy is a traditional theistic defense against the claim that, in view of the widespread suffering in the world, it is irrational to believe in the existence of the God of orthodox Western theism (an omnipotent, omniscient, and perfectly good creator of the natural universe).¹ Confronted with the atheistic argument from suffering to the non-existence of God (thus characterized), the soul-making theodicy argues as follows: "Freely chosen virtuous responses to suffering—e.g., freely chosen acts of steadfastness, charity and forbearance—are such that (1) it would be logically inconsistent to say that they occur and that suffering does not exist and (2) they are sufficiently valuable to outweigh the disvalue of the suffering which makes them possible. It follows from (2) that the fact that God made a world in which it is true both that suffering exists and free virtuous responses to it occur does not detract from His perfect goodness, unless He could have made a world in which there are free virtuous responses to suffering though there is no suffering.² But it follows from (1), in conjunction with the plausible claim that we do not really place restrictions on the power of an omnipotent being when we say that he cannot do what is logically inconsistent, that God could not have made such a world. Hence, the existence of suffering in our

world does not really show that God does not exist." In what follows, I shall consider some criticisms of this defense of orthodox theism and suggest some ways in which the soul-making theodicy might respond to them.

II

Soul-making theodicy holds that, all other things being equal, a world in which there is suffering and free virtuous responses to it is better than a world in which neither suffering nor free virtuous responses to it exists. This commits them to the view that free virtuous responses to suffering are intrinsically desirable, i.e., desirable regardless of whether they are means to a good end, rather than just instrumentally desirable, i.e., desirable because (as in the case of charity) they prevent or diminish suffering or (as in the case of steadfastness despite suffering) they promote good ends which suffering makes it difficult to bring about. (Plainly, it will not do to maintain that it is a good thing that suffering exists because it evokes free virtuous responses and that these are desirable for no other reason than that it is better than not that suffering should not exist or because it is good to bring about certain ends which could have been more easily achieved in the absence of the suffering.) Moreover, it is the contention of soul-making theodicy that free virtuous responses to suffering

¹ A recent, eloquent exposition of the soul-making theodicy can be found in John Hick, *Evil and the God of Love* (New York, 1966), pp. 207–400.

² It might be argued that this does not follow, since, instead of making possible free virtuous responses to suffering, God could have created, but did not create, things which are such that they do not have anything as disvaluable as suffering as a logically necessary condition of their existence but which are at least as valuable as virtuous responses to suffering; in which case He would have created a universe which is on balance better than our own. This objection presupposes that God is less than perfectly good unless He created the best of all possible universes, a thesis which is, I think, dubious. (See George Schlesinger, "The Problem of Evil and the Problem of Suffering," *American Philosophical Quarterly*, vol. 1 [1964], pp. 244–247.) But even if the soul-making theodicy grants its validity, he can still meet the present objection by maintaining either (1) that it may be, for all we know, that there is nothing which is as valuable as virtuous responses to suffering which does not have something which is at least as disvaluable as suffering as a logically necessary condition of its existence or (2) that it may be, for all we know, that there do exist somewhere in God's creation instances of something which is at least as valuable as virtuous responses to suffering and which does not have anything as disvaluable as suffering as a logically necessary condition of its existence and that God has not replaced virtuous responses to suffering with instances of this thing (i.e., created a larger number of instances of it while failing to make possible instances of virtuous responses to suffering) because a universe in which there is the maximum variety of types of valuable things is the best of all possible universes.

are sufficiently intrinsically desirable to outweigh the intrinsic disvalue of the suffering. But is this a plausible contention?

Dewey J. Hoitenga, Jr., has recently argued in the following way that it is not:

Bearing pain . . . is never a good state of affairs [even when the pain is, e.g., being nobly borne], for we are always morally obligated to prevent or remove it in so far as we are able. But . . . we are never obligated to frustrate a good state of affairs.³

This argument is designed to show, in effect, that the claim that free virtuous responses to suffering are intrinsically desirable commits the soul-making theodist to the view that we are not obliged to prevent suffering which is evoking a virtuous response—a view which no one, including the soul-making theodist, accepts. If the argument is sound, then the soul-making theodicy is deficient.

But is the argument sound? One reply which the soul-making theodist can make to it is as follows: Hoitenga's claim that we are never morally obliged to do away with a good state of affairs can be countered by two considerations. (1) It is far from clear that we could not be obliged to abolish a good state of affairs when in doing so we would *ipso facto* introduce one which is either better, or at least as good. Furthermore, (2) charitable acts of relieving suffering are at least as valuable as any other virtuous acts which the suffering may be evoking (as, e.g., a courageous response to the suffering), so that in relieving suffering which is already evoking a virtuous response, it may well be false that we are bringing about a world which contains less value than would otherwise have existed.

The atheist may, however, remain dissatisfied for the following reason. On the view of the soul-making theodist, our obligation to prevent suffering to which other people respond virtuously is in effect an obligation to prevent suffering which serves a certain valuable end, namely, that of evoking virtuous responses to the suffering other than our own virtuous response to it. But now some cases in which suffering serves a valuable end are cases in which we are not obliged to prevent it, and, indeed, *because* it serves a valuable end, we are in these cases obliged *not* to prevent it. Examples of such cases are the case in which someone is about to suffer the discomfort of having a diseased tooth removed and the case in which mild and brief suffering is being inflicted on a child for the

purpose of disciplining him. Now the atheist may wish to maintain that it is impossible for the soul-making theodist to point to a difference between the cases just mentioned and, e.g., the case in which somebody is bearing his suffering courageously, which would warrant him in saying that it is only in the former cases, not in the latter case, that we are obliged not to prevent suffering.

The objection is not, however, conclusive. The soul-making theodist can reply to it that the former cases are distinguished from the latter by the fact that the end which their suffering serves is that of avoiding future, greater suffering—that it is true in the former cases, but not in the latter, that in relieving present suffering we would in effect be causing more suffering than would exist if we did not relieve it. And the soul-making theodist can add (a) that it is plainly true that if in relieving a present instance of suffering one knows that he is *ipso facto* bringing about future, greater suffering which would otherwise not exist, then his act is an uncharitable act and hence is not intrinsically valuable, while (b) it is not plainly true that if in relieving suffering one is also doing away with a courageous response to it, then he is behaving in an uncharitable manner. Moreover, the soul-making theodist can contend that, indeed, the act of relieving suffering is, in the latter case, a charitable act, that charity is at least as valuable as courage; and that, hence, relieving suffering in the envisaged circumstances does not result in a loss of value. To this the atheist may wish to reply that if a courageous response to suffering really were sufficiently valuable to outweigh the disvalue of the suffering, then the act of relieving suffering would *not* count as a charitable act. But, though this may be true, it is, I think, neither intuitively nor demonstratively true, and, hence, the soul-making theodicy survives the present objection.

Another objection to the reply under discussion is that it follows from it that God cannot be exonerated on utilitarian grounds from the charge of moral reprehensibility for not preventing suffering. For (the objection runs) if human acts of preventing suffering which evokes virtuous responses are sufficiently intrinsically desirable so that in performing them we are introducing at least as much value into the world as would have existed if we had not performed them, then this ought to be true of God's acts of preventing suffering, i.e., it ought to be true that if God prevented all of the

³ Dewey J. Hoitenga, Jr., "Logic and the Problem of Evil," *American Philosophical Quarterly*, vol. 4 (1967), p. 117. See also H. J. McCloskey, "God and Evil," *The Philosophical Quarterly*, vol. 10 (1960), pp. 108-109.

suffering which He could prevent, the world would, because of the intrinsic desirability of God's acts of preventing suffering, contain no less value than in fact it contains at present.

This objection can also be met by the soul-making theodist. A person who causes another person to suffer with the end in view of giving himself the opportunity to relieve the suffering can be truly said to be creating the opportunity for himself to behave charitably with respect to the suffering only if it is true that he will subsequently regret his having caused it. However, God regrets none of His actions, since He is a perfectly good being and since a perfectly good being does not perform actions to which regret is an appropriate response. Hence, if God had created the conditions in which suffering can arise and then intervened to abolish it whenever it began to occur, then these interventions would not count as charitable (and therefore intrinsically desirable) acts. There is another way in which God could have prevented suffering, namely, by never having created it in the first place. But acts of refraining from producing suffering, unlike acts of relieving suffering with which one is confronted, do not count as charitable acts and, hence, if God had prevented suffering by having in the first place refrained from creating the conditions in which suffering can occur, this act, too, would not have been a charitable (and intrinsically desirable) one. It is of note that even if this latter contention is mistaken, there is another reply to the claim that the soul-making theodist is committed to the view that God's act of refraining from causing suffering would have been intrinsically desirable and that therefore He ought to have performed it: even if God's act of refraining from creating suffering *would* have been intrinsically desirable, it would not at least have been *as* valuable as the innumerable virtuous responses to suffering by finite persons which it would have made impossible. To this someone may be tempted to reply that because of its enormous effectiveness (in preventing all suffering) it would have been more valuable than all of the virtuous acts vis-à-vis suffering performed by finite creatures, since these latter, though they may diminish suffering, do not altogether abolish it. But to respond in this way would be either to forget that the soul-making theodist holds that charitable acts are intrinsically, as well as instrumentally, valuable or to

assume (what the soul-making theodist will wish to deny) that the degree of intrinsic desirability of a charitable act coincides exactly with its degree of instrumental desirability.

Still, in the end it must be conceded that the reply to Hoitenga which I have been discussing is not completely successful. It can be objected at this point that if the intrinsic value of God's refraining from introducing suffering into the world (granting that the act does have intrinsic value) would not have been as great as the intrinsic value of numerous human acts of relieving suffering, then, by parity of reasoning, the intrinsic value of one act of relieving suffering, performed by a single human being, is not as great as the intrinsic value of a larger number of acts of relieving the same suffering, performed by other human beings. Thus if I could know that a given instance of suffering would evoke numerous virtuous responses if and only if I did not abolish the suffering by a single charitable act, then I would not be obliged to perform the latter (anti-utilitarian) act—a conclusion which no one, including the soul-making theodist, would wish to accept.

Here the soul-making theodist needs to take a new tack. It will not do to argue merely that all morally obligatory acts of relieving suffering are sufficiently intrinsically desirable to balance or outweigh in value whatever valuable states of affairs the suffering may be giving rise to. Rather, he needs to maintain, as against Hoitenga, that it does not follow from the fact that one would, in relieving suffering, be diminishing the number of valuable states of affairs in the world that one is not obliged to relieve the suffering. He needs to contend, that is to say, that he has some deontological (as opposed to utilitarian) obligations to relieve suffering. Indeed, the soul-making theodist can be forced to this conclusion by considerations other than the one just mentioned. First, if you were responding courageously to your suffering and I were motivated to relieve it not by compassion or a sense of duty but, e.g., by an unseemly desire to enhance my own reputation (so that my act of relieving your suffering would not be a charitable act), I would nonetheless be obliged to relieve your suffering if I could do so.⁴ And, secondly, I am obliged to refrain from *causing* you to suffer even though I know that if I were to make you suffer then your suffering would evoke virtuous responses.

⁴ Cf. McCloskey, *op. cit.*, p. 108: "... some [scientific] discoveries [which have resulted in a decrease in suffering] have been due to positively unworthy motives and many other discoveries which have resulted in a lessening of the sufferings of mankind have been due to no higher motives than a scientist's desire to earn a reasonable wage."

There is no way for the soul-making theodist to account for these facts except by maintaining that we have, on the occasions referred to, anti-utilitarian obligations (a) to relieve suffering which would evoke virtuous responses and (b) not to cause it. For it is not open to the soul-making theodist to claim that my act of relieving your suffering would, when performed without compassion or conscientiousness, be a charitable, and, hence, intrinsically desirable, act (though, of course, it might be instrumentally desirable), nor can the soul-making theodist plausibly maintain that my refraining from causing you to suffer is an act of charity.

But is this new defense adequate? It will be objected to it that there is no difference between human beings and God which would account for its being the case that though the former have anti-utilitarian obligations of the kind just mentioned, God does not have these obligations. In answering this objection, I shall, for simplicity's sake, discuss just the claim that if I have anti-utilitarian obligations not to cause suffering which would evoke virtuous responses, then God does too. What is said about this matter will be applicable in an obvious way to the claim that since I might have the other sorts of anti-utilitarian obligations with respect to suffering mentioned above, it is likely that God has these other sorts of anti-utilitarian obligations also.

Suppose that God had an anti-utilitarian obligation never to actualize a possible instance of suffering such that the only good end which it would serve if actualized would be that of evoking at least one virtuous response to it. Then God would be obliged to act in such a way as to reduce enormously the number of virtuous responses to suffering which occur in our world. And, given that these responses are sufficiently intrinsically desirable to outweigh the disvalue of the suffering which evokes them, God's obligation to make them impossible would be a *vastly* anti-utilitarian one. On the

other hand, it does not follow that I have any vastly anti-utilitarian obligations from the fact that I have an anti-utilitarian obligation not to cause suffering such that the only good end which it would serve if it existed would be that of evoking at least one virtuous response to it. For (a) whether or not I cause suffering which evokes virtuous responses, the world will contain a great deal of suffering-cum-virtuous responses and (b) the number of virtuous responses which I can bring about by causing suffering is almost infinitesimal relative to the total number of virtuous responses to suffering which the world contains. It follows that it is true from the soul-making theodist's point of view that if the divine obligation which we are considering were, like my obligation regarding the causing of suffering, an obligation *never* to cause suffering just for the purpose of making virtuous response to it possible, then this divine obligation would be, unlike my obligation regarding the causing of suffering, a vastly anti-utilitarian one. And this is a plausible reason for maintaining that even though I have the obligation in question, God does not have it.⁵

But why suppose that if God had an obligation not to cause suffering just for the purpose of making virtuous responses to it possible, it would be an obligation not to actualize *any* possible instance of suffering just for this purpose? Why not say instead that God has an obligation to refrain from causing suffering just for the purpose of making virtuous responses to it possible right up to a point beyond which were He to *continue* to refrain from causing suffering just for that purpose, then (and only then) the total amount of virtuous response-evoking suffering which He had deliberately failed to cause would become so great that the act of deliberately refraining from causing all that virtuous response-evoking suffering would be vastly anti-utilitarian? To this the soul-making theodist can reply that God *does* have that obligation⁶ but that nobody can

⁵ "Can you bind the chains of the Pleiades or loose the cords of Orion? . . . Have you an arm like God and can you thunder with a voice like His?", *Job*, 38: 31; 40: 9.

⁶ It would, perhaps, be a bit more accurate for the soul-making theodist to argue here as follows: "There is no point at which God's refraining from causing virtuous response-evoking suffering would become vastly anti-utilitarian for the first time. Since God is omnipotent, any amount of virtuous response-evoking suffering which we might specify, no matter how enormous, is such that it will always be true of Him that He has refrained from causing that amount. Hence, no matter what quantity of virtuous response-evoking suffering we might specify as a candidate for virtuous response-evoking suffering which God is obliged to refrain from causing, it will always be true that there is *another* quantity which God has failed to cause such that His failure to cause both quantities combined is vastly anti-utilitarian. (A charge of moral reprehensibility against God cannot, of course, be based on this consideration, since God logically cannot avoid deliberately failing to make possible any amount of virtuous response-evoking suffering we may specify, no matter how enormous, and no one can be morally reprehensible for failing to do something when he cannot avoid failing to do it.)" The claim that any amount of virtuous response-evoking suffering which we might specify is an amount which God has deliberately failed to create presupposes that there is an indefinitely large number of possible people such that God foresees that, if He actualized them, they would freely respond virtuously to some possible instance of suffering, were He to actualize it. But it is difficult to see why this should not be the case.

prove that He has not fulfilled it. There is, after all, an indefinitely large amount of suffering which would, if actualized, have evoked virtuous responses which God has deliberately failed to cause, and no one can prove that this amount of suffering is not such that the act of refraining from causing not only this amount but still more suffering which would evoke virtuous responses would be a vastly anti-utilitarian one. Certainly it does not follow from the fact that *I* am not doing something which is vastly anti-utilitarian in failing to cause whatever virtuous response-evoking suffering *I* refrain from causing that God would not have been doing a vastly anti-utilitarian thing in failing to cause all of the virtuous response-evoking suffering which in fact He has refrained from causing *plus still more* (i.e., plus the virtuous response-evoking suffering which in fact exists in the world). For it is plainly false that *I* have refrained from actualizing anything like the amount of possible virtuous response-evoking suffering which God, if He exists, has refrained from actualizing.

Here the atheist may be tempted to respond with the following argument: "Though you (Dore) have not, strictly speaking, *refrained* from actualizing all the virtuous response-evoking suffering which God has refrained from actualizing, it is nonetheless true of you that you have *not* actualized the same amount of virtuous response-evoking suffering which God has refrained from actualizing, namely, all non-existent suffering such that it would, if actualized, have evoked virtuous responses. And if it is true that an obligation on God's part not to actualize all this suffering plus still more would be vastly anti-utilitarian, an obligation on your part not to actualize all this suffering plus all of the virtuous response-evoking suffering which you are able to actualize would also be vastly anti-utilitarian. But now either you are not obliged to refrain from causing suffering which would evoke virtuous responses if you brought it about or you *do* have the former obligation. It follows that if your

answer to the question why God does not have an obligation to refrain from causing the virtuous response-evoking suffering which exists in the world is sound, then *you* do not have an obligation to refrain from causing suffering which would evoke virtuous responses if it existed."

But the trouble with this argument is that my not actualizing all non-existent suffering which would have evoked virtuous responses if actualized is plainly not an *act* which I have performed, and so in speaking of an obligation of mine not to actualize all of this suffering, we are not speaking of something which I am obliged to do. And where we cannot speak of something which I am obliged to do, it is nonsensical to speak of an obligation of mine to do it. Hence, it is false that either I have no obligation to refrain from causing suffering which would evoke virtuous responses if I brought it about or I have an obligation not to actualize all the virtuous response-evoking suffering which I have not actualized plus all that I can actualize. My reasoning about God, on the other hand, does not suffer from this defect, since if God exists, it is true of Him that He has performed the act of refraining from actualizing whatever suffering does not exist.

Returning now to the claim that I have anti-utilitarian obligations not to cause suffering which would evoke virtuous responses, it must, of course, be admitted that this defense against Hoitenga fails if it is false that we ever have any anti-utilitarian obligations.⁷ Now I cannot, within the confines of this paper, discuss the very large question whether utilitarianism, taken as a general theory of the nature of all of our moral obligations, is true. Suffice it to say only that, at this point in the history of philosophy, it is far from certain that it is true, and that until the issue between deontologism and utilitarianism is firmly settled on behalf of the latter, the soul-making theodicy cannot be conclusively shown to be inadequate by Hoitenga's type of attack on it.

⁷ Someone might object that the obligation in question is not anti-utilitarian from a "rule-utilitarian" point of view, since refraining from causing suffering just for the purpose of making a virtuous response to it possible is an instance of following the general rule "Do not cause suffering except to prevent still greater suffering," and everyone's acting in accordance with this rule is more beneficent (value-producing or disvalue-preventing) than the universal following of any other relevant rule. Similarly, it is sometimes argued that my keeping a promise, even when my breaking it would give pleasure to someone and cause no harm, is not an anti-utilitarian act, since it is an instance of following a rule—viz., "Keep your promises unless doing so will cause great harm"—which is such that universal following of it is more beneficent than the following of any other relevant rule. But both of these views are mistaken. Surely the maximally beneficent rule in the former case is "Do not cause suffering unless in doing so you prevent still greater suffering or bring about some other end which is sufficiently valuable to outweigh the disvalue of the suffering," and the maximally beneficent rule in the latter case is "Do not break promises except to avoid great harm or when, in doing so, you bring into existence a valuable state of affairs which otherwise wouldn't have existed and, at the same time, cause no harm."

III

Another (very obvious) objection to the soul-making theodicy is that it is plainly true that much of the suffering which exists in the world does *not* evoke virtuous responses and that, even if the soul-making theodist is right in maintaining that God can be exonerated for permitting or causing those instances of suffering which do evoke virtuous responses, God is, if He exists, plainly morally reprehensible for causing or permitting those instances of suffering which are not redeemed by virtuous behavior.⁸

But suppose that God were to abolish or substantially reduce suffering to which we do not respond virtuously. (Call such suffering "apparently useless suffering.") Would the result really be a world about which the soul-making theodist must admit that it is a clear improvement over our own? In answer to this question I shall consider a number of imaginary worlds which differ from ours in that they either contain no apparently useless suffering or considerably less than exists in the real world, and I shall show reasons which the soul-making theodist might advance on behalf of the contention that these worlds are not plainly better than the real world.

The first imaginary world to be considered (call it W_1) is one in which whenever I am confronted with an instance of suffering with respect to which I choose not to be, e.g., charitable, God abolishes the suffering. In such a world I would come to know that there are two ways of relieving any suffering which would, in our world, call for a charitable response on my part: (1) take pains to relieve it (i.e., perform what count as charitable actions in our world) or (2) simply do nothing about the suffering.⁹ But in such a world I would know that there is no *point* in my using the first of

these two ways of relieving suffering. Hence, I would be neither morally praiseworthy for doing it nor morally blameworthy for failing to do it, and acts of relieving suffering the commission of which is not morally praiseworthy and the omission of which is not morally blameworthy are not, whatever else they may be, acts of charity. It follows that W_1 is a world in which charitable actions would be impossible, and it is open to the soul-making theodist to claim on this ground that W_1 is not clearly preferable to our own world.

At this point the atheist may wish to reply that since the soul-making theodist holds that taking pains to relieve suffering is an intrinsically desirable, as well as instrumentally desirable, act in our own world, he ought, in consistency, to admit that it would be an intrinsically desirable act in W_1 , even though it would not have the same instrumental desirability there, that therefore it would be morally praiseworthy in that world and that, hence, it would after all count as an act of charity there. But the soul-making theodist can reply that his view is that a necessary condition of its being true that taking pains to relieve suffering is an intrinsically desirable, charitable act is that doing so is the only means, or the most efficient means, of relieving the suffering which is available in the circumstances. And he can add that since taking pains to relieve suffering is frequently the only, or the most efficient, means of relieving suffering in our world but *ex hypothesi* is never that in W_1 , it is not at all inconsistent to maintain that taking pains to relieve suffering is frequently an intrinsically desirable, charitable act in our world but is never that in W_1 . Implicit in this reply is the thesis that the intrinsic desirability (and the charitable character) of the act of taking pains to relieve suffering is conditional upon the instrumental desirability of that act.¹⁰ But this would be dis-

⁸ Hick attempts to account for such instances of suffering on pp. 365-372, *op. cit.* He is not, I think, successful.

⁹ Unless God did one of the following things: (a) made me believe mistakenly that suffering to which I had responded uncharitably in the past had not vanished; (b) obliterated all recollection of anyone's ever having made uncharitable choices; (c) brought about a massive breakdown of my ability to reason inductively. However, the soul-making theodist can claim without extreme implausibility that a world in which none of these alternatives is realized is better than a world in which any of them is realized, even if the former contains more apparently useless suffering than the latter.

¹⁰ It would not, perhaps, be logically inconsistent to say that an act of taking pains to relieve suffering is an intrinsically desirable act and that it is not the only, or the most efficient, means of relieving the suffering which is available in the circumstances. And the atheist may ask, as a consequence, why God does not (a) replace our world with W_1 and (b) see to it that acts of taking pains to relieve suffering are intrinsically desirable even though they are not the only, or the most efficient, means of relieving suffering available. The answer is that though it may not be a logically necessary truth that acts of taking pains to relieve suffering are intrinsically desirable only when they are the only, or the most efficient, means of relieving the suffering available, this is nonetheless a valid value judgment which an omnipotent being can no more falsify (e.g., by decree) than he can falsify, e.g., the valid value judgment that it is wrong to inflict needless pain on people. Or if no being, for whom there is at least one value judgment which he cannot falsify, ought to be called "omnipotent," then the soul-making theodist can profess to be defending the existence of a being who is omniscient, perfectly good, and omnipotent in every respect except the present one.

tredding for the soul-making theodist only if it had the consequence that the intrinsic desirability of taking pains to relieve suffering was really nothing over and above the instrumental desirability of that act. And this is not so: to be conditional upon is not the same as to be identical with.

Another atheistic rejoinder is that charity would not be impossible in W_1 since doing nothing about the suffering of another person would come to count as an act of charity there, at least when it was accompanied by a feeling of compassion. But to this the soul-making theodist can reply that it is part of the definition of a charitable action that it involves effort on the part of the actor. And he can add that, even if this is not so, then at least it is the case that the charity which would be possible in W_1 is not nearly as intrinsically valuable as charity which does involve effort.

Thus far, in discussing W_1 , I have talked only about charity. But a similar point can be made about at least some other virtuous responses to suffering, e.g., steadfastness despite suffering. In a world in which God regularly abolished my suffering whenever I chose not to respond to it with steadfastness, I would learn¹¹ that I always have the following two alternatives with respect to instances of my suffering which would, in our world, call for perseverance on my part: (1) persevere in trying to achieve my end despite the suffering; (2) choose not to persevere while the suffering lasts and thereby make the suffering disappear. And since it would always be pointless for me to choose the former alternative, it would be neither morally praiseworthy for me to do so nor morally blameworthy for me to fail to do so. Hence, the choice to persevere would not be a moral choice, nor I, *qua* maker of it, a moral agent, and the range of moral actions which I could freely choose would be narrower in W_1 than in the real world. Could choices to persevere despite suffering properly be called "steadfast" in W_1 ? Perhaps not. But this is of minor importance. The important thing is that those choices would not be moral choices and that the soul-making theodist can claim with plausibility that this consideration counts as another reason for maintaining that W_1 is not clearly preferable to our world.

W_1 is a world in which God permits suffering to occur when I shall choose to respond immorally to it (until I learn that such choices will simply

result in the disappearance of the suffering, at which time they will no longer be immoral choices). But imagine a world in which it is true that when God (in His omniscience) knows that if a certain instance of suffering were to occur, then I would choose not to deal virtuously with it, He prevents the suffering from ever occurring in the first place, thus abolishing the opportunity for me to make any choice at all with respect to it. (Call this world " W_2 .") In such a world, I would not learn that when I, e.g., decide not to take pains to relieve somebody's suffering the suffering will simply disappear. Hence, I might *believe* (mistakenly) in W_2 that taking pains to relieve another person's suffering is an indispensable means to preventing his suffering, and, under these circumstances, my taking such pains could be morally praiseworthy and my failing to take them could be morally blameworthy; so that, though W_2 would contain no apparently useless suffering, it could contain acts of charity. (It could, for a similar reason, contain morally praiseworthy acts of steadfastness, etc.) In this respect W_2 is a better world than W_1 , and it may look as though it is preferable to our world. Why, then, has God not created W_2 instead of our world?

In response to this question it will be well to emphasize that it is an essential component of the soul-making theodicy that it is *freely chosen* virtuous actions with respect to suffering which are sufficiently intrinsically desirable to outweigh the disvalue of the suffering to which they are responses.¹² Virtuous actions with respect to suffering which are not freely chosen do not, on the soul-making theodist's view, have this characteristic. But now it is not clear that my choices to respond virtuously to suffering would be free choices in W_2 . For it looks as though it is a necessary condition of a choice of mine to do X at time t being freely made by me that I have the option of choosing at t *not* to do X . But when I choose, e.g., to take pains to relieve your suffering in W_2 it is false that I could at the same time choose *not* to do so, since if God had known that I would make this latter choice, he would have removed the opportunity for me to make it by never having permitted your suffering to start. (This is not the case in W_1 where, though it is impossible to choose to permit your suffering to continue unabated, I can, as an alternative to choosing to take pains to relieve it, choose to make

¹¹ With the qualification noted in footnote 9. Henceforth, this qualification will not be explicitly stated.

¹² A world of soul-making is, according to Hick, a world "in which moral beings may be fashioned, *through their own free responses*, into 'children of God'." *Op. cit.*, p. 293. (My italics.)

the suffering disappear by doing nothing about it.)

To this consideration the atheist may respond in one of two ways. Either he may refuse to accept the claim that my choices are not free unless there is some alternative choice open to me or he may accept this claim but deny that the question whether a choice meets the condition that there is some alternative choice open to the person who makes it is relevant to the question whether it, or the act which issues from it, is intrinsically desirable, maintaining instead that the question whether one is the *ultimate author* of one's own choice is the only question which has any bearing on whether the choice, or the act which issues from it, is intrinsically desirable. (There is nothing about the description of W_2 which rules out my being the ultimate author of my choices to respond virtuously to suffering there.) But even if the soul-making theodist grants the plausibility of one of these alternatives, he still has another reply to the claim that W_2 is preferable to our own world. He can argue that W_2 is not clearly better than the real world on the grounds (a) that in the real world our responding virtuously to suffering constitutes a real triumph over the real possibility of our behaving immorally, while in W_2 such a triumph is not open to us, since the possibility of moral defeat does not exist and (b) that actions which constitute a triumph over the possibility of moral defeat are at least *more* intrinsically desirable than actions which do not.

A third imaginary world which we need to consider is one in which (1) God permits suffering to begin even when He foresees that we shall not choose to respond virtuously to it (this distinguishes it from W_2) and (2) God intervenes to stop suffering to which we do not respond virtuously right up to the point beyond which it would no longer be possible to respond virtuously to suffering but not beyond that point. (This distinguishes it from W_1 .) Call this world W_3 . This world, like W_1 and W_2 , contains less apparently useless suffering than does the real world, and, in virtue of features (1) and (2), it is not defective in the ways that W_1 and W_2 are. Hence, the soul-making theodist needs to present a reason why God should not be held less than morally perfect for not creating W_3 instead of our world.

But perhaps this can be done. Let us look at W_3 a bit more carefully, and, for clarity, let us think about it in terms of our obligation to be charitable.

The precise point beyond which God's intervening to put an end to suffering to which we do not respond charitably would make charitable responses to suffering impossible is a point beyond which we would know that it is false that the likelihood of an instance of suffering continuing is greater if we do not respond charitably to it than if we do. Now to reach this point, God would have to intervene very frequently to abolish suffering to which we do not respond charitably (though, of course, He would not *always* intervene), and these frequent interventions would cause the odds that a person would continue to suffer when we neglected to be charitable to him to be considerably less than they are in our world. But now in a world in which the risk of your continuing to suffer in the event that I did not choose to be charitable to you was considerably less than it is in our world, my obligation to display charity to you when you are suffering would be substantially less stringent than it is at present¹³—rather more like my obligations to help old ladies across the street and not to drink more than two martinis (given that I have a normal tolerance for alcohol) than like, e.g., my obligations not to murder or maliciously destroy someone's reputation. And the soul-making theodist can base the following argument on this consideration: "A world in which we have a number of very stringent obligations is more desirable than a world in which we do not. It is important to some extent that we refrain from having more than two martinis but not nearly so important as that we fulfill our obligations not to commit murder. And it is important that some things should be thus extremely important, for it is a general rule that the greater the failure which would have resulted had one not triumphed, the more splendid is the victory, and this rule applies to resisting the temptation to do what is morally wrong. Now it is likely that reflection would show that all of our moral obligations regarding suffering are, like the obligation to be charitable, such that their stringency would be considerably reduced in W_3 . And, further, it appears to be the case that most of one's moral obligations have to do with suffering. Hence, W_3 is not after all a better world than our own."

IV

At this point, the following objection comes naturally to mind. It is one thing to show that a

¹³ This does not seem to be such that its contradictory is strictly speaking logically inconsistent, but the soul-making theodist can say that it is nonetheless a valid value judgment which even God cannot falsify.

world which has as little apparently useless suffering in it as W_3 would not be clearly preferable to our world and it is quite another to show that it would not be plainly better were *some* of the apparently useless suffering which exists in our world (though not so much as is missing in W_3) not to exist; and the soul-making theodist has yet to show the latter. Suppose that the world contained just a slightly smaller number of instances of apparently useless suffering than it in fact contains. Then plainly the kind of reduction of stringency of our obligations which would come about in W_3 would not obtain, and on what other grounds (the atheist may ask) could the soul-making theodist argue that the newly envisaged world would not be preferable to our own?

To this the soul-making theodist may wish to respond (1) that the degree of stringency of a person's obligations with respect to a given instance of suffering is a function of the intensity of that suffering and of the likelihood of the suffering going on for an appreciable period of time with that intensity given that he fails to fulfill whatever obligations he has with respect to it; (2) that were God to intervene to abolish or diminish some of the apparently useless suffering in the world, then He would bring about a state of affairs in which it would be at least slightly less likely, so far as at least one person was concerned, that a given instance of suffering would continue for a given duration with a given degree of intensity, and, hence, in which the degree of stringency of at least one of his obligations regarding suffering would be at least slightly reduced; and (3) that it cannot be proven that a world in which there is the sum total of the various degrees of stringency which in fact our obligations regarding suffering have is not slightly better than a world which is like it in all other respects except that the sum total of the various degrees of stringency of people's obligations is slightly less, despite the fact that the former world contains at least slightly more apparently useless suffering than does the latter.

An atheistic rejoinder to this is as follows: If it were true that God's abolishing one or a few instances of apparently useless suffering-cum-stringency would make this world slightly worse than it would be if He had not abolished that apparently useless suffering-cum-stringency, then this could only be because the lost stringency would be sufficiently valuable to outweigh the disvalue of the lost suffering. But plainly the loss of stringency which would result from the loss of one or a few instances

of apparently useless suffering would never be more than imperceptibly small, and an imperceptibly small amount of stringency could never be sufficiently valuable to outweigh the disvalue of the suffering with which it was connected.

But the trouble with this argument is that, if it is sound, it proves that it is clearly true that God, if He exists, is morally obliged to abolish *each* instance of apparently useless suffering which exists in our world and, hence, *all* of the stringency which attaches to our obligations regarding suffering. For if the argument is sound, then God would never arrive at a point in the process of abolishing apparently useless suffering beyond which the stringency lost due to the abolition of one or a few more instances of apparently useless suffering would be more than imperceptible and, hence, He would never arrive at a point such that were He to go beyond it by abolishing still more apparently useless suffering, He would bring about a world which was slightly worse than the preceding one. But it is not *plainly* true that a world with *some* apparently useless suffering-cum-stringency is not preferable to a world in which there is neither any apparently useless suffering nor any stringency attaching to our obligations regarding suffering. Moreover, anyone who agrees that this is not plainly true must also agree that there *would* come a point in the process of abolishing apparently useless suffering-cum-stringency such that any reduction of apparently useless suffering-cum-stringency beyond that point would make for a slightly less perfect world than the one which went before. Must he also admit that at this point any further stringency which was lost as the result of the loss of one or a few instances of apparently useless suffering would not be imperceptibly small or that, if it would be, it would nonetheless be sufficiently valuable to outweigh the disvalue of the lost suffering? The soul-making theodist does not have to answer this question. (He can simply decline to accept the invitation to step into this Zeno-type trap.) All that he needs to say here is that, whatever the explanation may be of the fact that the world lying immediately beyond the envisaged point would be slightly less perfect than the one immediately preceding that point, it would *ex hypothesi* be less perfect. And he will wish to add that it may be, for all anybody can prove to the contrary, that God does not intervene to abolish the apparently useless suffering in our world because our world is in fact at a point such that worlds which were exactly like it in all respects except that they

contained less apparently useless suffering-cum-stringency would be at least slightly less perfect.

Here the atheist may wish to argue in turn that if the soul-making theodist really believes that were God to abolish even one instance of apparently useless suffering, He would diminish the stringency of at least one obligation regarding suffering to an undesirable extent, then the soul-making theodist is committed to the view that he, too, would be reducing stringency to an undesirable extent in doing away with any apparently useless suffering¹⁴ and, hence, that he is never obliged to be charitable. But even if the soul-making theodist grants (what possibly may be doubted) that it is plainly true that if God would be reducing the stringency of at least one obligation regarding suffering by abolishing some apparently useless suffering, then human acts of abolishing suffering would have the same effect, the soul-making theodist can reply that many acts of relieving apparently useless suffering are, on his view, sufficiently valuable to balance or outweigh in value the stringency which they do away with, and, hence, that he is not committed to the view that in acting to relieve apparently useless suffering he is always bringing about a world which is slightly inferior to the way it was before he acted. This reply does not hold good for acts of relieving apparently useless suffering which are motivated neither by compassion nor a sense of duty and are therefore neither charitable nor intrinsically desirable, but the soul-making theodist can maintain that he is not committed to holding that we are not obliged to perform these acts but only to holding that our obligation to perform them is anti-utilitarian in nature. "But then," the atheist will say, "since you admit that some of our obligations to abolish apparently useless suffering are anti-utilitarian, you must explain why God does not also have an anti-utilitarian obligation to abolish apparently useless suffering." To this the soul-making theodist can respond either that it may be that God already abolishes so much apparently useless suffering-cum-stringency (e.g., in answer to prayer) that an obligation on His part to abolish this much and still more would be *vastly* anti-utilitarian or (setting this aside) that it may be that God has deliberately failed to cause so much apparently useless suffering-cum-stringency that an obligation on His part both to fail to

cause that much *and* to abolish some that He has caused would be vastly anti-utilitarian. And he can add that my obligation to relieve apparently useless suffering even when I do not feel compassion and am not motivated by a sense of duty is not itself a vastly anti-utilitarian obligation since I have not deliberately failed to cause anything like the amount of apparently useless suffering-cum-stringency which God has failed to cause and, given that God abolishes apparently useless suffering-cum-stringency (e.g., in answer to prayer), then I have not abolished nearly so much as God has abolished.

Still, even though it cannot be shown in the manner just considered that there is something wrong with the claim that there is no apparently useless suffering in the world which is such that its abolition would not lower at least slightly the stringency of at least one obligation regarding suffering, this claim may be doubted. If we are not inclined to doubt it, this is probably because we are thinking about apparently useless suffering which is such that some human being has an obligation to respond virtuously to it and, hence, which is such that if it does not evoke a virtuous response from some human being then it evokes an immoral response. Now if all apparently useless suffering were like this, then the question why God does not prevent it would be the same as the question why God does not prevent suffering with respect to which human beings respond immorally. And the answer to this question may well be that were God to do so, then He would abolish or reduce the stringency of human obligations regarding suffering. But now in fact there are some instances of apparently useless suffering which do not create any obligations on the part of human beings. For, though it may be doubted that any (or anyway much) human suffering is such that it does not create an obligation on the sufferer's part (e.g., an obligation to try to bear his suffering with at least a modicum of grace or courage) or on the part of people other than the sufferer (e.g., an obligation to try to relieve the suffering),¹⁵ at least much of the suffering of lower animals does not create obligations on the part of human beings, because, for one reason, it takes place on occasions when no human beings are present. Moreover, this suffering is not connected with any stringency which human obligations possess: it is plainly false

¹⁴ Since suffering which one relieves for the right motive is not apparently useless (in my sense of "apparently useless"), it would be more accurate (though also more awkward) to say that he would be reducing stringency in doing away with any suffering which *would* be apparently useless *if* he did not eliminate it.

¹⁵ But imagine a man dying alone on a desert island who is no longer able to try to bear his suffering courageously.

that were God to relieve or prevent such suffering, then He would reduce or abolish the stringency of any of our obligations. (Any inference from the fact—if it were a fact—that the suffering of lower animals is short-lived or non-intense or non-existent to the conclusion that human beings are not apt to suffer intensely or long when we fail to respond virtuously to their suffering would, in view of what we in fact observe about human suffering, be wildly fallacious.) And, finally, it will not do to suggest that though the suffering of lower animals is not connected with the stringency of any human obligations, it is nonetheless connected with the stringency of the obligations of lower animals themselves: since lower animals are not (presumably) endowed with free will, they have no moral obligations and, hence, are not obliged to respond virtuously to their own suffering or to that of their fellow brutes.

But now does it follow from the fact that the suffering of lower animals is not connected with the stringency of any human (or lower animal) obligations that it is in no way related to the stringency of *any* obligations at all? The soul-making theodist can deny this with some plausibility. He can claim that it cannot be disproven that it is for the most part true¹⁶ that when the suffering of lower animals is not caused by human beings, it is brought about by the free choices of moral agents other than human beings,¹⁷ and he can give the following reason¹⁸ why God allows these agents to have the power to cause lower

animals to suffer: The disposition of a person to display extreme cruelty is often the result of his having frequently chosen in the past to be uncharitable, unyielding, overly-competitive, and so on. Moreover, our obligations to practise the virtues of charity, forbearance, etc., are more stringent than they would be if a person who failed to practise these virtues did not run the risk of making himself exceedingly cruel. Now not even God could have brought it about that it is as urgent as it is in fact for a person to fulfill his obligations to be charitable, forbearing, etc., if people did not run the risk of becoming exceedingly cruel as a result of failing to do so¹⁹ or if God prevented people from making uncharitable, unyielding, etc., choices or prevented people from manifesting their disposition to behave cruelly, once they had developed that disposition as a result of making those choices. And it may be that God makes it likely that we shall develop a disposition to be cruel when we choose to be uncharitable, etc., and that He does not prevent us from choosing to be uncharitable or from manifesting our disposition to be cruel, once we have developed that disposition, because it is intrinsically desirable that it is as urgent as in fact it is that we fulfill our obligations to be charitable, etc., and sufficiently intrinsically desirable so that this urgency (this stringency of our obligations) outweighs the disvalue of the suffering which results from cruelty.²⁰ If this is true, then if non-human moral agents inflict suffering on lower animals as a result of their having become cruel

¹⁶ I say "for the most part" because it may be that the suffering of pets can be accounted for by the soul-making theodist in terms of human obligations to which it gives use. In future I shall, for convenience, drop this qualifier.

¹⁷ Alvin Plantinga points, in a slightly different context, to the possibility that non-human spirits may be responsible for some of the world's evils. See Plantinga's *God and Other Minds* (Ithaca, N.Y., 1967), pp. 149ff.

¹⁸ An alternative reason is simply that the envisaged non-human moral agents having this power is a logically necessary condition of an intrinsically desirable state of affairs, namely, their also having the power freely to choose to *refrain* from causing the suffering of lower animals. It has been maintained earlier (in Sect. II) that God's act of refraining from causing suffering would not have been intrinsically desirable. But perhaps the soul-making theodist can add that acts of refraining from causing suffering *are* intrinsically desirable *if* they are the outcome of moral struggle (of which God is presumably incapable). Nonetheless, it is, I think, implausible that choosing to refrain from being cruel to lower animals, even after moral struggle, is by itself sufficiently valuable to outweigh the disvalue of their suffering.

¹⁹ This is not quite true. Our obligations to be charitable, etc., would have been as stringent as they are so long as *some* dire consequence of being uncharitable, etc., comparable but alternative to the likelihood of eventual cruelty envisaged above, took place. But, at any rate, God needed *some* such consequence to see to it that the aforementioned obligations are as stringent as they are (see footnotes 10 and 13) and the likelihood of cruelty seems at least as good as any.

²⁰ The soul-making theodist will probably want to add that though the mere existence of stringent obligations *which are never in fact fulfilled* is not sufficiently valuable to outweigh the disvalue of the suffering of lower animals, it may be that non-human moral agents do in fact more often than not freely choose to fulfill their obligations to be charitable, etc. (Moreover, he will probably want to say that God does not prevent the beings under discussion from failing to fulfill their obligations to be charitable, etc., nor just because this would make the envisaged stringency impossible, but also because it would render their choices to be charitable, etc., *unfree* (see *W₂*) or, at any rate, not moral choices (see *W₁*) and because virtuous choices must, to be intrinsically desirable, be freely made, moral choices.) The fact that the soul-making theodist may well want to say that the non-human moral agents which we are envisaging perform a balance of virtuous over wicked acts seems to me a good reason for not following Plantinga (*op. cit.*) in calling them "devils." In view of the traditional meaning of "devil," the use of the term in the present context may lose more in terms of misleadingness than it gains in terms of abrasive anti-modishness.

in the envisaged way, we can understand why God should permit this to be so.

It is of note that on this view the *actual* suffering of brutes is not a necessary condition of some of the obligations of non-human moral agents being as stringent as they are, for these beings might freely choose to *refrain* from making those choices which result in their disposition to behave cruelly, or, once having attained this disposition, they might freely choose not to make it manifest. But if their obligations are to be as stringent as they are, then it must be *likely* that they will become disposed to behave cruelly *if* they choose to be uncharitable, etc., and God cannot prevent them from actualizing that disposition, once they have attained it, *if* they freely choose to do so. And the soul-making theodist will wish to claim that unfortunately some non-human moral agents *have* chosen to be uncharitable, etc., and that, having thus become disposed to be cruel, they choose to behave in a cruel manner to lower animals.²¹ Or, at any rate, he can maintain with plausibility that it cannot be proven that this is not so, and, hence, that it cannot be proven that the suffering of lower animals is such that God could prevent it while failing to abolish the stringency of somebody's obligations.

At this point the atheist may wish to argue that even if there is no way of demonstrating that the envisaged non-human moral agents do not exist and that it is not a good thing that their obligations are as stringent as they are, it is nonetheless plain that the soul-making theodist does not really believe that this is the case, for if he did believe it,

then he would suppose that, e.g., in relieving the suffering of a horse who has broken his leg, he would be to some extent diminishing the stringency of at least one obligation of at least one non-human moral agent and, hence, he would not believe himself morally obliged to relieve the horse's suffering. But there are a number of ways in which the soul-making theodist can reply to this objection. First, he can maintain that those relatively few instances of animal suffering which give rise to human obligations are the work of God rather than of non-human moral agents. And, setting this aside, he can point out that it is his view that most acts of relieving animal suffering are sufficiently intrinsically desirable to compensate for the value of the stringency which they eliminate. The soul-making theodist must admit that this answer does not cover acts of relieving animal suffering which are motivated neither by compassion nor a sense of duty and, hence, are on his view neither charitable nor intrinsically desirable, but he can claim that we have an anti-utilitarian obligation to perform these acts. And with regard to the question why God does not also have an anti-utilitarian obligation to relieve instances of animal suffering, the soul-making theodist can argue, as we have envisaged him doing in similar contexts earlier, that it may be that God does not have such an obligation because, in view of the enormous amount of suffering-cum-stringency which He has deliberately failed to create, the obligation would, if He had it, be vastly anti-utilitarian in nature.

Vanderbilt University

Received July 11, 1969

²¹ But why do they choose to be cruel to lower animals, instead of to men? Perhaps God does not permit them to do the latter. But what motive do they have for torturing lower animals? Are they insane? Perhaps what motivates their cruelty is their desire to frustrate God's wish that all men should know of his existence by making it appear as if no omnipotent and omniscient being could be perfectly good.

IV. DISPUTED EVALUATIONS

FRANCIS E. SPARSHOTT

HUME has often been cited in support of a strange thesis. According to this thesis, conventional books on ethics start by discussing what *is* the case and then slide imperceptibly into talking about what *ought* to be the case, with no logical justification for the transition from one copula to the other.¹ The citers generally take Hume to be denouncing the error of putting "ought" into the conclusion of an argument in whose premisses only "is" figures,² and take this to show how easy it is to confuse facts with values and how clearly Hume saw that they must be distinguished. The thesis is a strange one because, in the first place, Hume gives no examples of the condemned maneuver and his followers seldom supply the deficiency;³ thus one cannot tell what fallacy would be committed. In the second place, what seems the most likely form for the condemned transition would not be fallacious. This would be a syllogism with two premisses with "is" and a conclusion with "ought": arson is forbidden by Jehovah; burning down the White House is arson; ergo, one ought not to burn the White House down. But what is supposed to be wrong with this argument? In any one-semester logic course one used to learn two things. First, one must not expect to find the copula "is" used in real life wherever a syllogistic argument is employed: to show the logical form of the argument, one must

twist sentences around. Thus one might have expected to read, not "Arson is forbidden by Jehovah," but "Jehovah forbids arson," and the argument would have been none the worse. But the same procedure requires one to transpose "One ought not to burn the White House down" into "Burning the White House down is something one ought not to do." The other thing one learned in that one-semester course was that in everyday life syllogistic arguments usually appear in enthymematic form. No one ever bothers to state both premisses, so that to show the form of the argument one has to write in the omitted premiss. In the example before us, the missing premiss is obviously "Everything that Jehovah forbids is something one ought not to do." And the argument now stands revealed by the most elementary procedures as two impeccable syllogisms in Barbara. Thus if this is the sort of argument that Hume and his clique had in mind, and I don't know what else it can have been, one has to conclude that either none of them had taken even a one-semester course in logic or none of them were thinking about what they were saying. The latter alternative is likely: since they had before their minds no actual instances of the alleged pattern of argument they had no occasion to ask whether any fallacy would really have been committed.

¹ David Hume, *Treatise on Human Nature*, ed. by L. A. Selby-Bigge (Oxford, 1888), pp. 469-470. "Ought" is there said to be, or to stand for, a "new relation or affirmation" that requires explanation, and of which one cannot say how it "can be a deduction from others, which are entirely different from it." Numerous modern citations are listed by Nicholas Capaldi, "Hume's Rejection of 'Ought' as a Moral Category," *The Journal of Philosophy*, vol. 63 (1966), pp. 126-137, note 1. Capaldi argues that the "other" relations must be causal relations, not that expressed by the copula "is"; but the immediate context certainly suggests that the latter is meant.

² A. C. MacIntyre, "Hume on 'Is' and 'Ought,'" *Hume*, ed. by V. C. Chappell (New York, 1966), pp. 240-264, points out that "deduction" in Hume is more likely to mean "inference" than "entailment," so that Hume may be taken as merely "asserting that the question of how the factual basis of morality is related to morality is a crucial logical issue." But if this means that moral argumentation is non-deductive, one feels bound to say that deductive arguments are simply those in which the ground of inference from reason to conclusion is made fully explicit. Most people who give examples of allegedly non-deductive moral arguments merely cite conversational exchanges in which the nature of the logical connection between ground and consequence remains wholly obscure.

³ P. H. Nowell-Smith, *Ethics* (Baltimore, 1954), p. 37, note 2, quotes Bishop Robert Cecil Mortimer, *Christian Ethics* (London and New York, 1950), section 7: "The first foundation is the doctrine of God the Creator. God made us and all the world. Because of that He has an absolute claim on our obedience. We do not exist in our own right, but only as His creatures, who ought therefore to do and be what He desires"; and adds that the argument requires the additional premiss that a creature ought to obey his creator. In fact, Mortimer's argument seems ambiguous: it is not clear whether he is arguing that I ought to take my dependence as a reason for obeying, or that the fact of my dependence is a reason why I ought to obey. In either case, what Nowell-Smith represents as a suppressed premiss seems rather to be what Mortimer is chiefly concerned to affirm.

Granted, if someone did put forward something like our example in the belief that it was a properly and completely formulated argument, he would have been wrong. But how could anyone make such a mistake? The incompleteness and informality cannot be missed by anyone capable of formulating the notions of validity and logical form, and hence of claiming formal propriety for his argument.

Let us suppose that someone did commit the Humean blunder, whatever it is, and in some fashion that we need not specify mixed *oughts* with *ises*. Would he then be confusing facts with values? To say that would be to use the terms "facts" and "values" in strained senses. No doubt he would be confusing descriptions with prescriptions, or assertions with injunctions, but what have injunctions to do with values? The language of obligation and its logic are, *prima facie*, far removed from the language of evaluation and its logic, and the bridges and tunnels that philosophers build between them are rickety and leaky affairs.⁴ If we wish to erect a barrier between sentences with "is" and sentences with "ought," it is not at all obvious on which side of that barrier sentences with "good" or "value" will fall, or even whether they will all fall on the same side.

Facts and values are certainly not the same. They are so different that they could never be confused. One can start a sentence with the phrase "The fact is that . . .," but it makes no sense to start one with the phrase "The value is that. . . ." But this is no help to those who want to dissociate goodness from factuality, because it puts goodness on the wrong side of the fence. It makes sense to say that it is a fact that Islay Mist is a good scotch; it makes no sense whatever to say that it is a value that Islay Mist is a good scotch. Values are either things, prized objects, or else standards by which things are appraised; and facts, whether defined as correlates of propositions or in whatever

way, are neither things nor standards. It follows immediately that facts can be ascertained and values cannot; but this fact of ontology or logical grammar has not the smallest tendency to show that value judgments are not assertions of fact.

What this shows is perhaps only that the contrast between "facts" and "values" is posed in misleading terms. *Of course*, if we define a value as "any object of any interest,"⁵ then that any given thing is a value is merely a fact about it like any other fact. But not all that is value is valuable, and to call something valuable is not to give any information about it but to evaluate it. Granted, if I tell you that my first edition of Ephiphanius is not so valuable as I hoped it was going to be when I bought it, this is informative, but we can brush that aside. Even so, the mere distinction between the valued and the valuable does not suffice to push attributions of valuability into the domains of subjectivity and expressive utterance.

Given four kinds of linguistic doings, giving information about something, saying how good it is or evaluating it, saying what one feels about it, and evincing one's feelings toward it, what are the affinities between them? The last two tend to coalesce, since saying how one feels is often the easiest and directest way of showing one's feelings. To show a girl one loves her, one can say "I love you." On the other hand, saying or showing how one feels about something is quite different from giving information about it, except in the merely irritating sense that if I love a girl she is loved by me and that she is loved by me is a fact about her of which she or anyone else can be informed. But saying or showing how one feels about something is just as easily told from saying how good it is or evaluating it. People who are required to write letters of recommendation or similar reports are expected to keep their feelings out of them. If there is an exception to be made, it is in favor of those

⁴ See my *Enquiry into Goodness and Related Concepts* (Toronto, 1958), ch. 8. I should now take a firmer and clearer line on the matter than I did then. R. M. Hare, reviewing the book in *The Philosophical Quarterly*, vol. 10 (1960), pp. 372-374, finds a contradiction. I wrote (p. 135) that "To speak of a need is to speak of a deficiency which is 'really' there for everybody to recognize; and to imply that everyone 'ought' to recognize it"; and Hare, calling this a definition of "need," and pointing out that I define "good" in terms of "need," insists that I have thus defined "good" in terms of "ought" and hence (since I elsewhere accept Hare's prescriptivist account of "ought") committed myself to Hare's own brand of prescriptivism. These are desperate measures. The sentence quoted is in no sense a definition of "need," and does not even occur in the section of my book where needs are chiefly discussed. The sentence itself precludes a purely prescriptivist account of needs, since presumably what is spoken of takes precedence over what is implied. Had I offered a definition of "good" (which I didn't), and had "need" figured in that definition (which it wouldn't), I should still not have defined "need" as anything like "that which ought to be taken into account" in any respect. Hare seems to suppose that because I said (p. 247) that I was prepared to accept "ought" as an indefinable term I must somehow be defining all other terms in terms of it. But the supposition is unwarranted. *Prima facie*, no contrast could be stronger than that between the opted and the imposed, matters of choice and matters of obligation. The contrast has been forcibly stated, in a different frame of reference, by Julius Kovesi, *Moral Notions* (London, 1967), pp. 149-160.

⁵ R. B. Perry, *General Theory of Value* (Cambridge, Mass., 1926), p. 115.

whose expert knowledge is so much a part of them that their feelings about things of a certain kind are a reliable index of their merits. We often do make, and are entitled to make, this sort of assumption in the areas of our competence; but that does not mean either that our feelings are what we are talking about or that the purport of our report is to express our feelings. It is important that, as Aristotle affirmed and Dewey repeated,⁶ evaluations and appraisals are not only matters where expertise is relevant but are in the proper domain of expertise. What stands in the way of an evaluation is often ignorance, and it is just when an evaluation is required that one calls on an expert. It is surely from this sort of consideration that reflection about evaluation has to begin. And if so, any treatment of evaluation that makes knowledge only incidentally or accidentally relevant is ruled out from the start as wrong-headed, even if logically viable. Similarly, any account that makes discussion of values impossible or inexplicable is merely perverse. Values are pre-eminently discussable: what else is there to discuss? There is no point in arguing about feelings. And one only argues about facts if either they cannot be checked or it cannot be checked that they have been checked. Evaluating, then, is clearly different from expressing feelings; but, because it is discussable, it is perhaps also different from imparting information. However, it is possible that evaluating something is not different from giving one sort of information about it: namely, information about how good it is.

The words "evaluate" and "evaluation" are themselves used in a variety of ways. In many contexts, to evaluate something (for example, an influence) is merely to determine its extent or magnitude; in other contexts, to determine its probable market price. In so far as these senses merely reduce evaluation to giving specific information of a special sort, they do not concern us.⁷

On the other hand, some philosophers use "evaluation" and its cognates as terms to cover all utterances that contain such terms as "good," which they call "value words." But this usage is also inconvenient for us, since in calling something "good" one may be doing any of a wide variety of things.⁸ We are here concerned only with those applications of "good" that can meaningfully and intelligently be contradicted, challenged, supported and argued about. In this sense, to call something good is a limiting case of saying *how* good it is, and saying how good something is is what we want to mean by evaluation.

Saying that highways 50 and 9 afforded in 1965 a better route from Thistle town to Orangeville than highways 7 and 10 is pretty clearly neither showing how one feels about them nor giving positive information about the number of traffic lights, type of surface, etc., but either something in between or something different. This, I take it, is nowadays common ground. If there is a live question, it is what sort of positive account one is to give of such utterances. One recently popular form of account found in them both a sort of informativeness and something else, such as commendation. Evaluations were said to combine descriptive with emotive or prescriptive meaning, in varying proportions and with varying priorities. But Stevenson's and Hare's fundamental versions of this dichotomy, brilliant and instructive as they were, were too simple.⁹ The ingredients did not cohere. Despite the precautions of their authors, the logic of these accounts made evaluations look too arbitrary in relation to their supporting reasons, so that reasoning from facts to an evaluative conclusion became at best causally efficacious rather than logically cogent. There is much to be said for a more traditional mode of treatment, one that was perhaps more congenial to wielders of languages that disposed of gerundive verb forms, according to which calling something good is

⁶ Aristotle, *Nicomachean Ethics* 1096a27-34; John Dewey, *Theory of Valuation* (Chicago, 1939), pp. 5, 22, and *passim*.

⁷ They do, however, concern us closely to the considerable extent that they go beyond information into speculation. An expert appraiser does not say what the price is but what it *would* be, in accordance with what he conceives to be the state of the market, which he infers from actual transactions. He estimates what factors have affected those transactions, the weight that was likely given them, the likely changes since their time, the degree of analogy in relevant respects between them and the case submitted to him. In the typical case the appraisal is not subject to confirmation, because it is performed *in lieu of* the market transaction that alone could have confirmed it.

⁸ Although I was careful to call "good" "a philosopher's dummy," it was perhaps a mistake in method to present my case primarily in terms of that word and only secondarily in terms of evaluation. It required me to make the inelegant move of treating many uses of "good" as "secondary" or "derivative." For an account of "Evaluation" see now Jon Wheatley in the *American Philosophical Quarterly*, vol. 5 (1968), pp. 199-205; for tentative moves toward a position like his, see my *Enquiry*, *op. cit.*, pp. 64, 98, 212, 251-252, 293.

⁹ Charles L. Stevenson, *Ethics and Language* (New Haven, 1944), *Facts and Values* (New Haven, 1963); R. M. Hare, *The Language of Morals* (Oxford, 1952), *Freedom and Reason* (Oxford, 1963).

primarily asserting that it *is* good; and to be good is to be, roughly, choiceworthy, an appropriate object of choice or desire in some implied circumstances. No doubt to call something good would sometimes, though not always, also implicitly convey some quite definite information about the properties that made it choiceworthy; and no doubt it would also, on occasion, be to perform some such illocutionary act as commending or recommending. It might even, in the circumstances touched on before, be to express one's feelings about it. But the primary task of an analysis of goodness would be to explicate what I have called choiceworthiness. It was to this task that I addressed myself some 15 years ago in the central part of *An Enquiry Into Goodness*.¹⁰ What I want to do now is to recapitulate the leading features of the account I gave there and to draw out certain implications of it that passed unnoticed.

The analysis was organized around a formula: "To say that *X* is good is to say that it (1) is (2) such as to (3) satisfy (5) the (4) wants of the (7) persons (6) concerned." (I have inserted numbers to mark the loci and sequence of seven types of disagreement, to be discussed below.) The formula was meant to evade the embarrassing fork, that a naturalistic meta-ethics commits the naturalistic fallacy and a non-naturalistic one has somehow to smuggle nature in through the back door. The formula evaded the latter prong by coming very close to what I tried to show was the proper pattern of analysis for all property concepts. It evaded the former prong by a certain allusiveness, which would have been more obvious if it had been cast in direct speech: "To say '*X* is good' is to say '*X* is such as to &c'." Anyone who says that something satisfies "the persons concerned" is clearly relying on certain prior judgments, such as who the persons concerned actually are. But these judgments form no part of what he is saying. Psychologically, he may have in mind certain persons, of whom he is thinking under a particular

description. But logically he is referring to whoever falls under that description. He is not therefore equating goodness with the satisfaction of certain *specified* desires in the way that objectors to "naturalism" find fallacious.¹¹ I was careful not to write "'Good' means 'such as to satisfy &c'," a formula that would have implied that the question of who was concerned had somehow been taken care of, and would have reduced to a familiar sort of equation of excellence with a variety of satisfactoriness selected and elaborated by the author. My use of a more complex formula was meant to prescind from any such selecting and, by contrast, to accommodate all possible differences of opinion as to what is good and what is not. What the formula is meant to do is precisely to provide a means of articulating such differences themselves. To use Dewey's distinction, it is the kind of theory that starts from "value" as a verb rather than "value" as a noun.¹²

The formula asserts that two forms of utterance are equivalent, that is, mutually substitutable. But under what conditions? *Salva veritate*, no doubt, but the intended emphasis is more pragmatic than semantic. The thesis is that evaluations are disputed and defended as if they had been cast in the longer form, and that that form shows the range of disagreements in evaluations. If someone says that a thing is such as to satisfy the wants of the persons concerned and someone else denies it, there are about seven things they may in fact be disagreeing about.¹³ I shall list seven, and then consider what kinds of disagreement they are.

First, two people who respectively affirm and deny that something is good may be disagreeing about whether it has the "good-making" properties which, as both agree, would if present make the thing such as to satisfy those concerned. In this case, the subject of disagreement is an ordinarily empirical question, and its evaluative context makes no difference.

In all the other modes of disagreement, the dispu-

¹⁰ Chs. 4-7 and 9. Because of my immaturity, showing itself partly in changes of aim and partly in excessive readiness to be swayed by older and wiser heads, the book suffered from a confusion of purpose; I here give myself credit for clearer and more consistent intentions than I had.

¹¹ A simplified version of the fundamental situation may be diagrammed as follows:

<i>X</i> is good	contradicts	<i>X</i> is not good
<i>X</i> satisfies those concerned	contradicts	<i>X</i> does not satisfy those concerned
Those concerned are <i>A, B, C</i>	contradicts	Those concerned are (not <i>A, B, C</i> but) <i>D, E, F</i>
<i>X</i> satisfies <i>A, B, C</i>	does not contradict	<i>X</i> does not satisfy <i>D, E, F</i>

¹² John Dewey, *op. cit.*, p. 4.

¹³ The points may be made anything from 6 to 9, depending on certain arbitrary choices in classifying possibilities.

tants may agree about what properties (in the above straightforward sense) the thing has, but disagree about whether those properties render it such as to satisfy &c. We shall see that it is at least not obvious that any of these is an ordinarily empirical question.

The first of these modes, and thus the second mode of disagreement altogether, is about whether a thing which admittedly gives satisfaction in the situation currently in question is itself *such as* to give satisfaction or merely happens to have done the trick for now. At issue here are the source and reliability or repeatability of the satisfaction.

Third, disagreements may arise out of differing notions about what constitutes satisfaction, whether for example it is getting what one craved or getting something that allays craving.

Fourth come disagreements about wants. "Wants" was chosen as a term broad enough to cover needs as well as desires and requests. Even if it were always a plain matter of fact that someone desired or had asked for such-and-such, what he needed might still be a matter of opinion as to what standards were to be applied; and there could be further questions as to whether agreed needs should take precedence over agreed desires in determining what someone "wanted."

Fifth, even if these sorts of disagreements about wants did not arise, there could still be argument as to which wants were *the* wants: the weak demonstrative "the" makes pre-emptive claims that may well be disputable.

Sixth, people may disagree as to who is concerned. There is more than one way of being concerned in an affair. Besides those participating, there may be others who do not like to be told that it is none of their business. And among those who are admittedly concerned some again may be ruled out by the pre-emptive force of the definite article as having an interest too marginal to be considered.

Seventh and last, if we take a "person" to be a

being conceived capable of entering into personal and moral relations with other persons, including the person who is doing the conceiving, then there may be disagreement about who is a person.¹⁴

The interpretation of my analytic formula obviously hinges on the characterization of these kinds of disagreement. If they can all be settled by straightforwardly empirical methods, by looking and seeing and measuring and counting, the formula turns out to be indirectly naturalistic after all, even if it manages to avoid at least some of the pitfalls that have been called "the naturalistic fallacy."¹⁵ If on the other hand all the disagreements reduce to divergences of taste or attitude, then the formula turns out to be covertly emotivist, though its emotivism is in some measure recollected in tranquillity. If some approximate to one kind and others to the other, the formula may reduce to a combination of naturalistic and emotivist elements such as emotivists claim for emotivism. With this in mind, we briefly examine them in reverse order.

First, then, how is it decided who is to be regarded as a person? There seems to be more than one way of choosing criteria of capacity for moral and personal relationships. One can choose the active criterion, that someone behaves as a moral being, recognizing obligations and so on; or one may recognize him as passively having certain rights, so that other people treat him in a moral way. To have rights, it may suffice that one should be thought capable of feeling, capable of being affected as a sentient being by the actions of others; but to have obligations one needs at least enough intelligence to understand social situations and moral ideas. In deciding on such a basis who is a person, one may then be deciding whether the values of feeling and passivity are paramount, or whether regard should be had only to those who can be members of a functioning community; and on either basis one would have to decide what degree of supposed sentience or actual participation

¹⁴ Do these seven exhaust the possible kinds of disagreement? Exhaustiveness is never provable unless the field to be classified is specified either extensionally or by the principle of classification itself, neither of which is the case here. But the very notion of a dispute about wants may be enough to yield the following. There must be (4) a real want of some kind, somehow established, that is (7) a want of some want and (3) a want *for* something. On the basis of these three, all disputable, may be established a configuration of a wanted thing, relative to a single want. Given more than one want, the relative configurations may be incompatible, in which case harmony must be established among wants of (5) one want or (6) different wanters. Thus may be established an agreed configuration of a desideratum. Then if anything it may be asked whether it matches this configuration, either (1) at some time or (2) all the time or in general.

¹⁵ This is true at least in so far as "persons concerned" cannot be given a naturalistic interpretation, since there is nothing in the formula to guarantee that empirical methods of identifying them are employed, even if such methods are in principle available. This may be a technicality rather than a point of substance, but people who use the vocabulary of "fallacies" are committed to technicalities.

constituted personhood. The boundaries of the moral community may be variously drawn.¹⁶ Deciding who is a person, either in a specific context or in general, is thus deciding which of two sorts of factor is to be allowed more (or exclusive) weight, and within each kind at what level qualifications are to be set. But whatever differences may then arise are limited by the strictly controlling factors that limit the possibilities of actually engaging in social action or being affected by it. Weightings may vary, but what is weighed does not, and there are limits beyond which weightings require elaborate defense if they can be defended at all.

Since the "wants" of our formula include desires and requests as well as needs, the materials of the formula suffice to support an alternative argument on the requirements for personhood.¹⁷ A person must be a being capable of wanting, and such that his wants are in some circumstances entitled to consideration, such too that they *can be taken* into consideration. That is, people must be able either to find out what he wants or to imagine and postulate what he may or must want. Two requirements seem to be thus imposed. First, the being must be enough like ourselves that we can sympathize with him, at least in imagination. Second, communication with the being must be possible, so that his wants can be made known. These requirements come close to specifying what Aristotle meant by a "political animal," a being capable of rationally formulating common ends of action with others; and they have jointly seemed to make "person" synonymous with "human being." On this showing, who is a person would be almost a straightforward matter of fact. Almost, but not quite; for those who equate humanity with personhood can refrain from inconvenient expansion of the moral community by deciding that some races that one might otherwise have called human have only recently descended from the trees and belong to a hominid species of quasi-ape. Setting that aside

as a piece of obvious special pleading, it does seem that all of us men can be distinguished from all of those others the beasts, and any man can learn any other man's language and ways as he can never learn the "language" and ways of geese or even of chimpanzees. And yet, as we know, the equation of man with person has not always seemed compelling. One may not care to learn a "barbarous" language, or may even refuse to admit that it is a proper language at all. Or one may just decide not to talk or listen to the barbarophones, and make the essential division between one's tribe and all else. On the other hand, theists regard their deities as persons, and one may tell one's troubles to God or the bees or the yams,¹⁸ and listen to what they tell one in one's dreams. And yet this does not reduce the question of who is a person to one of caprice. The requirements of personhood are as stringent on this basis as on the other, and if we do not *find* that they are fulfilled we must *decree* that they are so. The considerations brought forward do not demand that personhood and humanity be equated, but they are such that those who would interpret personhood otherwise must pretend that something happens that is not observed to happen, or adopt canons of evidence here such as they would elsewhere reject, or interpret phenomena in arbitrary or strained ways, or at best rely on merely conventional and adventitious relationships rather than inescapable likenesses and differences. Thus to make humanity the basis of personhood may well be, as J. S. Mill said it was,¹⁹ the mark of a sophisticated and critical intelligence as opposed to a mentality in thrall to the folkways, even if it remains true that most people are quite unsophisticated and very uncritical, and that all of us are sometimes uncritical or find critical reasons for not giving up what we once uncritically held. From all of this, three conclusions can be drawn. First, the decision as to who is a person is neither fully determined by empirical considerations nor

¹⁶ Even within a single society they need not be drawn in a consistent way. On the boundaries of personality there may be beings recognized by the society as having rights but no duties (e.g. furry mammals) or duties but no rights (e.g. slaves). Duties and rights are correlative only in the sense that if someone has a right someone else must have a duty, and vice versa.

¹⁷ There is a third way of setting out requirements for personhood. The presupposition of the whole argument is that the wants in question are those of which the wantor may be aware. The arguments in the text extend personhood to those with whom we do form a moral community, and those with whom we could form a moral community, that is, those with whom conscious wants are or can be communicated. But a less stringent requirement is that a person is at least a being capable of consciously wanting, that is, (a) a living thing, (b) a thinking thing, and (c) in some measure autonomous (or the notion "I want" could hardly occur to it). But if we concede that such a being exists, how can we say that its wants are *in principle* unknowable to us? And if we could know them, could we find a good reason for ignoring them?

¹⁸ I mentioned in my book (p. 166, note 46) a Melanesian group whose concept of personality allegedly extends to yams, but not to white men. Nor is this absurd. The vegetables growing in my garden are obviously in closer symbiosis with myself than human foreigners are.

¹⁹ Cf. J. S. Mill, *Utilitarianism*, ch. 3.

purely capricious. Second, it makes sense to *try to determine* who *should* be regarded as a person, rather than just find out who is so regarded; and third, if one does try to decide, what one does is less like looking into one's heart to see how one feels, on the one hand, or trying to establish what is the case, on the other, than it is like trying to determine what *weight* is to be given to factors whose *relevance* imposes itself.

The question who is concerned in a given affair, or kind of affair, has some affinity with the question of who is a person, inasmuch as personhood can be equated with capacity for being concerned in affairs in general, but of course it is not the same question. In many contexts, what is said shows that those concerned are a reference group. A good wine is one of a sort that meets the relevant demands of those who concern themselves with wines;²⁰ and, in general, expressions of the form "a good so-and-so" do not invoke persons concerned with the so-and-so in question on a particular occasion, but the class of so-and-so buffs. Setting aside such "functional" evaluations as unproblematic, one can say that in any affair there are some persons whose concern can be denied only by denying their personhood. To deny that the agent or patient of an action is concerned in it is in effect to deny that he is truly agent or patient: if he has interests, one must admit that they are affected. But there may be other parties to a transaction whose interest is more oblique or controversial. People often argue about whether certain affairs are "any concern of" the police, or society at large, or God; and one may be condemned as a busybody or as apathetic for differing from one's neighbors as to whether some affair is or is not one's concern. Some people have interests that are certainly affected, but only remotely, or slightly, or indirectly; others may be in such a position that their interests will perhaps be seriously affected but perhaps will not be touched at all. Some are concerned because they concern

themselves, make something their concern; some are concerned because their very knowledge of an affair offends their susceptibilities; some people declare that matters in which they could intervene are "no concern of theirs" and they do not wish to *become* involved.²¹

Having made up our minds in a given case about which if any of the doubtful cases are genuine concerns, we still have to decide who are "the" persons concerned: that is, the ones we are going to take into account (no matter how many others there may be) and thus treat as the only ones. The characteristic use of the definite article with a plural noun is typically significant for my argument: it does not mean all and it does not merely mean some; it means "the ones we are going to attend to"—all, or most, of the ones that matter. Wherever "the" is used with a plural noun there is plenty of room for disagreement, and what such disagreements turn on is relevance together with importance. How close, direct, and inescapable must a concern be to be taken into consideration; and how do these and other factors affect the weight to be given it? Such a question suggests a calculus of concerns; but, as with the Benthamite calculus of pleasures, the term "calculus" is misleading inasmuch as there is no basis for a rational quantification and computation.²² The question is not really one for calculating any more than it is one of brute fact or caprice. It is one that calls for, precisely, weighing up and estimating, and for differential attention.

In turning from concerns to wants, we meet an important complication due to the difference between needs and desires. It seems at first that the question of what a person desires is a purely factual one: to find out what someone would like, ask him. Desires can be converted to requests, and requests can be recorded. On the other hand, to say what someone needs is to pass judgment on his condition, and the one judging must be assumed responsible for his judgment. Every mature individual is

²⁰ Cf. my *Enquiry*, *op. cit.*, section 6.1632. This pattern of analysis is precluded by insistence on rendering "the good *X*" as "the *X* which is good" (cf. Jerrold Katz, *The Philosophy of Language* [New York, 1966], p. 290). But the only reason for this insistence seems to be that it facilitates a certain set of analytic maneuvers in general grammar (cf. Noam Chomsky, *Syntactic Structures* (The Hague, 1965), p. 72, and one does not see why one should impoverish one's own exegesis merely to accommodate the convenience of another discipline. The same move seems to lie behind Paul Ziff's otherwise inexplicably jejune conclusion in *Semantic Analysis* (Ithaca, 1960) that "'good' in English means answering to certain interests" (p. 247).

²¹ Certain East African farmers, told recently that their farming methods would render their land barren within a generation, replied: "Our sons must face their own problems." Similar attitudes cause the widespread pollution of the lakes, streams, and air of North America.

²² Arbitrary quantification is of course possible: cf. John C. Hall, "Quantity of Pleasure," *Proceedings of the Aristotelian Society*, vol. 67 (1966-67), pp. 35-52. In fact the practice of relying on precise computations of sums in terms of units and scales arbitrarily established is too common.

usually allowed to be the sole judge of what he desires or "would like," but we do not usually concede to him or to any one authority the right to lay down definitively what he needs. To say that something is needed is to say that there is, by some relevant standard, a deficiency to be supplied,²³ and both the relevance of standards and their application seem to lie in the public domain. In fact, to claim a need is often to make an open appeal to the public domain. But within that domain no authority can be precisely located, and the appellant must in the end rely on his own best judgment.

We seem to have found a quite sharp split between factual ascriptions of desire and non-factual ascriptions of need. But the appearance may be deceptive. By asking a man what he wants we certainly establish what he requests, but even requests may be misunderstood. And desires leave more room for doubt. Desiring is not like having an itch, it is wanting *something*, and every desire may be formulated as a desire "that something should take place." But what a desirer designates to himself or others as that which should take place may be unworthy of that designation—since it has not yet happened, it may turn out when it does happen not to be at all what was looked for. And even if it does prove to be what it was expected to be, it may not yield the expected gratification. One can therefore argue that a man may be mistaken about what his desires are in at least two ways, and thus cannot after all be the final authority on his desires. But to this it can be replied that desires are not thereby shown to be unfactual,

but rather that there are three plain facts involved: that a man wants 'X' in the guise of 'X'; that what he wants in the guise of 'X' has really the character not of X but of Y; and that his quest for Y in the guise of 'X' will lead to disappointment if what he gets is either X or Y rather than Z, which would turn out to be "what he really wanted all the time."²⁴ If that is the right way to put it, what we have uncovered is not something unfactlike in desires but further complexities in the notion of "want," a term that covers a multiplicity of facts. Attribution of need, on the other hand, remains neither mere recognition of fact nor pure invention: to be plausible, such attribution must be based on objective considerations that seem, if not to demand it, at least to suggest it strongly. Claims not so based are brushed aside. The mechanic who tells me that my car needs new shocks is neither merely describing the state of my present shocks nor expressing his feelings, but, on the basis of a description, appealing to a norm. A car-owner may reject the appeal, but to reject the norm would betray stark ignorance or blind folly.

Even if we now say that there is something factlike about at least some alleged needs, and something perhaps unfactlike about desires, the two are not confusable; one does not always need what one desires or desire what one needs. Thus deciding what "wants" are *the* wants requires us to balance against each other two sorts of things that are not directly comparable. Yet in choosing or evaluating we may have to decide between them,²⁵ which is why I used the generic term "wants" for both. It could therefore be thought that in preferences

²³ Or what would be a deficiency if not guarded against. For the necessity of this correction, see David Braybrooke, "Let Needs Diminish that Preferences May Prosper," *Studies in Moral Philosophy, American Philosophical Quarterly Monograph No. 1* (1968), pp. 85–107, note 23, on "course of life needs."

²⁴ Herman wants to marry Nathalie. Nathalie, unknown to Herman, wears falsies, and has a nasty side to her character that she is equally careful to hide from her suitor. The "Nathalie" projected by Herman's desire is thus not Nathalie as she really is, although it is undoubtedly Nathalie and no one else that Herman wants to marry. However, even if Nathalie were all Herman's fancy paints her, marriage to her would still disappoint him, since his conscious yearning masks a deep unconscious craving for a relationship with a handsome busboy like Max.

²⁵ It may be argued that the concept of need is pre-emptive, so that to call anything a need is to say already that it has precedence over all desires (cf. Braybrooke, *op. cit.*, p. 87). But I should claim that this is true only in a limited context. First, it is not true that I should allow what *someone else* says is my need to have automatic precedence over my desires. Secondly, the precedence holds even for one person only if the standard by which the deficiency (or potential deficiency) that generates the need is assessed is a standard directly relevant to the desire for what is supererogatory in relation to it. Thus I cannot deny that the fact that I need a salary of at least \$10,000 next year to avoid radical changes in my way of life is more important than the fact that I should like a salary of \$25,000 so that I can live in yet greater comfort. But the fact that I need a haircut does not drive me to a barber if I would like to go to a movie and cannot do both, nor do I see why it should.

It is in the specialized field of public welfare that the pre-emptiveness of needs is most at home. It seems natural, if funds are to be distributed, for the responsible agency to establish a level below which none must fall whether they want to or not and whether they apply for relief or not, to specify this level in concrete terms of specific goods and services, and to meet the "need" to reach this level before trying to get people what they happen to like. But should we expect the recipient to accept the bureaucrat's scale of values? Braybrooke (*op. cit.*) writes in general as if there were only one set of agreed needs current in any society, and as if the public welfare context were the only relevant one. But these suppositions are unwarranted.

between needs and desires we find for the first time a kind of disagreement that depends not on divergence of weightings but on plain difference in choice or attitude. But that is not so in all cases, if ever. Whether we forego something we long for in favor of something we need will depend on how badly we crave the one and need the other and what we crave and need them for. In weighing need against desire no less than in weighing need against need and desire against desire, it seems that we are trying to determine what the situation calls for rather than ascertaining the facts or merely assessing our feelings, and that in such determination we are weighing up factors whose relevance imposes itself.

Next, satisfaction. Some questions about satisfactions are merely alternative versions of questions about desires, and therefore need no separate treatment.²⁶ But there is another kind of disagreement, to which an analogue can be found in arguing about needs. To say that someone is satisfied may be to say that he has had "enough," and how much is enough is often debatable. More generally, being satisfied is a matter of degree: one can be more or less satisfied, and concede something to be satisfactory merely in despair of anything better.²⁷ In all such cases, debates on whether something is satisfactory turn on weightings, estimatings, seeing whether something will do. And seeing whether something will do is neither determining its nature nor seeing how one feels about it, nor a blend of these, but a clearly distinguishable third beside them.

Next, being *such as* to satisfy. One may concede that a thing has given or is giving satisfaction (in whatever sense) and still deny that it is *such as* to do so, for to say that a thing is *such as* to do something is to impute its doing so primarily to its own nature and properties,²⁸ rather than to something special or adventitious in its temporary

conditions and surroundings that makes its doing what it has on this occasion done exceptional.²⁹ It is also to say that it can be counted on to do so, not necessarily always, but most of the time and as a rule. Judgments that *X* is *such as* to do *T* impute causal dispositions, and are thus largely factual; but for two reasons they cannot in every case be fully supported by any routine or empirical procedures. First, there may be no agreement on how reliable performance must be to sustain imputation of a disposition. Second, such imputations carry an element of indefinite prediction of future performance. Thus a weighing up of probabilities is involved. But this element of weighing up is doubtless less prominent here than in some other cases we have examined.

Finally, as we said at first, questions of whether things actually have what would admittedly be good-making properties are on the same footing as other empirical questions. They therefore pose no special problems for us, though perhaps we should remind ourselves that whether something is square, or red, or rough, or measures $2\frac{1}{4}$ by $2\frac{1}{4}$ inches, is not always something that can be settled by examination or measurement alone without deciding where borderlines are to lie and whether they have been crossed. Most of the time, we establish our tolerances as we go. Even such yes-or-no questions as whether Caesar has crossed the Rubicon may be easier to answer at some seasons of the year than at others.

Let these then be the kinds of disagreement in evaluation that the formula admits. It has been my contention that these and no others are the kinds of disagreement that do arise, but I do not see how this can be demonstrated and I do not wish to argue the point here. If it be granted that evaluations do resolve into these components, what are the consequences for evaluation in general?

First, a contrast between fact and value has not

²⁶ For example, in so far as disagreements about what is satisfactory hinge on whether what is needed to fulfill desire is what a person thinks he would like or what he would later on acknowledge to be what he had wanted all the time, or on whether satisfaction is achieved by causing a desire to cease or only by providing that which was desired, they are precisely what we were discussing in discussing desires.

²⁷ In academic grading systems, the grade of *C* is often given the verbal equivalent of "satisfactory," but the satisfaction is less than hearty.

²⁸ There is an important difference in the implications of being "*such as* to" do something as applied to people (at least in moral contexts) and to things, answering to different grounds of constancy in performance. With things, this is their nature and properties; with people, rather their will to do so. There is also a lesser ambiguity, in things, between actually having certain enabling properties, and belonging to a kind which tends to perform in a certain way.

²⁹ There is here the apparent conceptual difficulty that surroundings of some kind are always present and relevant: *X* can only be *such as* to do *T* in circumstances *Z*. But the difficulty is factitious. At least in all cases where the distinction between satisfying and being *such as* to satisfy is applicable, in specifying the wants and the persons concerned we are stipulating standard types of situation, within which and relative to which evaluation is performed. What fits in a particular set of such surroundings will not necessarily fit in all such sets, though of course what fits all such sets will fit every such set.

shown up anywhere. Although only one of the seven kinds of disagreement was about facts, none of the others had a character that contrasted in any striking way with factual disagreements, and all were more like differences about fact than like differences in feeling or (to use Stevenson's formulation) in attitude. They differed from factual disagreements in the extent to which they involved an element of *weighing up*: not the degree of approximation to a paradigm, but rather the importance of factors whose relevance was not reasonably deniable. It is above all to the paramount importance for value theory of these three concepts of weighing up, importance, and relevance, that my argument leads.

What is relevant to any problem or affair is almost but not quite determined by how things are and go. Knowledge of relevance is largely knowledge of causal connections and matters of fact. What makes the presence of someone in the next apartment relevant to my decision to play the piano is the body of relevant facts—that the walls are not soundproof, the man is not deaf, he is sick, he is distressed by piano playing as I play it. I can of course still *say* that these facts are all irrelevant to my decision, but that is of no consequence. Anyone can say anything; what matters is rather what one can expect to get accepted. Like many such statements, this declaration of mine will be no more than an *ex parte* refusal to take things into account, and will convince no one. I cannot make something irrelevant by deciding to ignore it. "X made no difference to A's decision to do Y" does not mean the same as "X was irrelevant to A's decision to do Y." But it does not seem easy to say in general terms what relevance is, what it is to have something to do with something.³⁰ This is after all a very familiar notion, and one might not

expect to find any simpler or clearer terms to explicate it in. Perhaps we might try the following. A proposition *p* is relevant to A's decision *d* if it can be shown *how* the truth of *p* could affect *d* without any special information about A's thought-processes. But this suggestion has the weakness that the notion of "what can be shown" covertly contains the notion of rationality. Crazy people can "show" what is not there to be shown, stubborn people cannot be shown what is there to be shown, gullible people can be "shown" what is not there to be shown, and stupid people can neither show nor be shown what is there to be shown. And the element of evaluation inherent in this notion of rationality opens the door to those who, like Stevenson, would re-introduce on this level the whole debate about evaluations. To pronounce on a thing's relevance, they will say, is only to declare one's attitude toward taking it into account.³¹ The case for this sort of move is sometimes thought to be strengthened by a supposed distinction between the "factually relevant" (what causally affects or is affected by an event) and the "morally relevant" (what normatively should be taken into account by a judgment or decision).³² But this is not a distinction, discovered within the concept of relevance, that happens to support the Stevensonian analysis; rather, it replaces the concept of relevance by two other concepts for which no use has been established and which are framed with no other purpose than to meet the requirements of that analysis itself. The considerations we adduced before should suffice to make such maneuvers even less convincing in this context than on their original home ground, where they have few defenders left; but the case cannot be said to be closed.

The concept of importance occupies a key

³⁰ The concept of relevance is treated at some length, though within a restricted context, in my *Concept of Criticism* (Oxford, 1967), ch. 15. I there suggested that *X* is relevant to *Y* if it can be connected up with it in the appropriate way, the type of connection appropriate being not arbitrary but specified by the type of inquiry or undertaking in hand. Cases of doubtful relevance would be those where the connection either could not be made out to the satisfaction of all, or where the connection was doubtfully of the appropriate kind; borderline cases would be those where the connection was far-fetched but evident. But none of these concepts, either by themselves or in combination, seem either less vague or easier to grasp than the explicandum itself.

³¹ Wayne Leys seems to lay himself open to Stevensonian interpretation when he writes that "A judgment of relevance expresses a belief about the *worthiness* of a proposition for investigation or deliberation in relation to a given problem or on a given occasion." ("Irrelevance as a Philosophical Problem of Our Time," *Memorias del XIII Congreso Internacional de Filosofía* IV [1963], p. 185; my italics). But this is deceptive. To pronounce something worthy of investigation is not so much to express an attitude toward it as to pronounce on the likelihood of the investigation turning something up: a fundamental point akin to that made by Plato in the *Theaetetus*, 178–179.

³² Cf. Donald C. Emmons, "Moral Relevance," *Ethics*, vol. 77 (1966–67), pp. 224–228. Mr. Emmons writes (p. 225) that "We can assess the purpose of any proposed action from a normative standpoint, and until we have done so the moral relevance of a particular factual consideration is indeterminate." This seems to open the door to those self-serving *ex-parte* declarations of irrelevance stigmatized in the text.

position beside that of relevance. No more than relevance can importance be plausibly made out to be a matter of how anyone feels. I take it that to be important is not to attract attention or to be the object of an attitude, but actually to make a difference to people, to affect the courses of their lives. How I feel about a thing, what attitude I take toward it, may affect how much difference it makes to me, but hardly how much difference it makes in general; and calling something important to oneself or some other particular person is so far from being the same as calling it important *sans phrase* that it is usually used as a device for conceding its unimportance. If it mattered I would not need to say that it mattered *to me*. Magnitudes of changes in the courses of lives cannot be assigned precise quantities, but they can certainly be compared, and one can meaningfully discuss not only which of two things is the more important, but how important one thing is.³³

Alongside determinations of relevance, it seems that *weighings-up* or *estimations* of importance must take a central place in any realistic theory of evaluation.³⁴ "Estimation" here cannot mean, as it often does, guessing at a quantity that might have been and perhaps later will be precisely determined; it means rather saying how seriously something must be taken in matters where no possibility of non-arbitrary quantification and calculation enters. Plato was no doubt the first to contrast questions of measurement and enumeration with questions of more and less, which he thought of as inherently indeterminate concepts.³⁵ His mathematicizing bias led him to regard questions of this latter sort as unamenable to rational treatment and hence to relegate them to the sphere of emotion. The prejudice remains, and to claim theoretical importance for such notions as weighing up, deciding whether things will pass muster, and the like, is still to court accusations of anti-intellectualism, hence irrationality, hence advocacy of emotionalism. But, like other Platonic prejudices, this is hardly more than a quaint survival. If one considers

the kind of thinking demanded by such practical problems as driving through rush-hour traffic, it seems undeniable that it involves ignoring some movements in one's vicinity,³⁶ attending marginally to some others, concentrating on a few, taking into account some theoretically possible eventualities and not others, and making more or less allowance for those one does take into account. None of these factors are calculable in practice, but the impossibility of calculating does not make any aspect of the process inexplicable, unjustifiable, or in any sense irrational. Still less does it mean that the rush-hour driver must or does make emotion his guide. In fact, the effect of emotion, in any ordinary sense of that term, is to impair the driver's responses and the quality of his driving. The joint impossibility of calculation and irrelevance of emotion that characterizes practical problems is precisely what makes it both possible and necessary for lawyers in cases involving the exercise of care and discretion to appeal to the "common sense" of the disinterested bystander, the "man on the Clapham omnibus," and what made it both necessary and possible for Aristotle throughout his *Ethics* to invoke the hypothetical judgment of the "prudent man." But the necessarily severe limitation on formal treatment of such practical judgment does seem to have the consequence that, as Aristotle remarked and his example suggests, there is not much that can usefully be said about it. Hence it tends to be ignored in theoretical discussions. But this neglect leads to a serious misrepresentation of practical reasoning, since it inevitably comes to be supposed that what cannot figure largely in theory cannot matter greatly in practice.

Absence of an agreed unit of measure and technique of measurement means that such estimations and evaluations as we are considering cannot be rendered immune to challenge and in this sense must remain matters of opinion. But not all opinions are equally sensible and some are silly. Some are obvious, some are far-fetched. Some require special pleading and ad hoc hypotheses, some do

³³ Judgments of importance are of course often made with exclusive reference to a matter in hand, so that irrelevance becomes as it were a limiting case of unimportance; but they are not necessarily so made. See the next note.

³⁴ We have spoken as if one established *whether* something was relevant, but *how* important it was. But one can also discuss how relevant it is. In practice, this is likely to come to the same thing as discussing how important it is but in theory how relevant something is should depend on how close, direct, and few are the steps by which the connection is made, rather than how much difference would be made if the connection were conceded. If a matter can (as presumably it can) be highly important but totally irrelevant to a given issue, it can presumably be highly important though marginally relevant.

³⁵ Plato, *Philebus* 24a-25b.

³⁶ Learning to drive is largely a matter of becoming so habituated to driving situations that one does not notice the innumerable movements on and around the road that are irrelevant to one's own actions. But this dismissal, though automatic, is a rational judgment of relevance and is made correctly by every good driver.

not. The fact that everyone is silly some of the time and some are silly most of the time does not show that the silly and the sensible are indistinguishable: on the contrary, if they could not be distinguished the remark would be incomprehensible. On the other hand, even the silliest evaluation is an evaluation, and can be sympathetically understood by anyone to whom the necessary connections and weightings are explained.

As Stevenson says that the end of disputes about values is to get people to revise their attitudes,³⁷ we must say that their end is to get people to revise their weightings. And such revisions are for the most part brought about by adducing facts or establishing causal connections.

In sketching this pattern of analysis for evaluations, I am not suggesting that all rival patterns are untenable. It should indeed be obvious that any pattern that has recommended itself to competent philosophers must accommodate all relevant phenomena without intolerable strain. Preferences

among patterns of analysis depend on what cases one takes to be typical, at what points one is prepared to introduce qualifications and subsidiary hypotheses, what problems one considers it most important to meet head-on.³⁸ I have chosen a pattern that softens the contrast between "facts" and "values," between constating and evaluating. Since everyone agrees that reasons for evaluations as for actions are most often given by citing facts, it seemed better to adopt a pattern that showed from the start how facts can be relevant and suggested what facts would be relevant, rather than begin by sundering values from facts and then lamenting, like a child who has dismembered the family clock, that one cannot reunite what one has separated. But of course it remains open to any proponent of mid-century orthodoxy to show that my analysis is a mere halfway house and that more diligent explication would reveal, within my bricks of importance, relevance, and weighing up, straws of the pure evaluative essence.³⁹

Victoria College, Toronto

Received June 5, 1969

³⁷ Charles L. Stevenson, *Ethics and Language*, (New Haven, Yale University Press, 1944) *passim* (e.g., pp. 13, 17, 140).

³⁸ There is here an obvious danger of being caught in a vicious circle: in explaining what he means by claiming superiority for his own analysis of superiority, every analyst must on pain of bad faith rely on his own analysis. Stevenson can only urge us to approve of his; I must claim that mine is such as to satisfy the wants of those who want analyses.

³⁹ It is of this essence that Mr. Kovesi writes, parodying Marx, that it is "the universal, independently constituted value of all things which has therefore deprived the whole world, both the world of man and nature, of its value. The 'evaluative element' is the alienated essence of man's work and his being" (*op. cit.*, p. 143).

V. EGALITARIANISM AS A DESCRIPTIVE CONCEPT

FELIX E. OPPENHEIM

LIKE "democracy" or "freedom," the term "equality" has a laudatory connotation. Hence the tendency to apply it to those, and only those, institutions or policies which one wishes to commend, and to qualify as inegalitarian those of which one disapproves. The question then arises: Is saying that, say, a sales tax or a graduated income tax is egalitarian like maintaining that one or the other is equitable? If so, persons with different views as to whether a certain policy is equitable are bound to disagree as to whether it is egalitarian, and communication is likely to break down. If, on the other hand, it were possible to set down descriptive criteria of equality, it would become possible to discuss in a meaningful way whether egalitarianism in general is just, or whether egalitarian principles of a particular kind are desirable. It seems, therefore, worthwhile trying to explicate the concept of egalitarianism so that it yields criteria which are not only empirical, but also general. Then we can ascertain whether any given rule—sales tax or graduated income tax, universal military training or student deferment, to each according to his work or need—is egalitarian or inegalitarian, regardless of whether it is equitable or just or desirable on some other grounds.

I. "EQUALITY" IN THE EXPRESSION TO BE DEFINED

First, let us determine the expression we want to define. Here we must distinguish. "Equality" can be predicated either of certain characteristics of persons, or of distributions made by one actor to at least two others, or of rules stipulating how such distributions are to be made. "Equality" in the first two meanings presents no problem from the point of view of our topic, and we shall be mainly concerned with equality as a property of rules of distribution.

¹ To this point, cf. Hugo A. Bedau, "Egalitarianism and the Idea of Equality" in J. Roland Pennock and John W. Chapman (eds.), *Nomos IX; Equality* (New York, Atherton Press, 1967), pp. 3-27; esp. p. 7.

² Thomas Hobbes, *Leviathan*, ch. XIII.

³ Ernest Nagel, *The Structure of Science* (New York, Harcourt, Brace & World, Inc., 1961), pp. 492-494.

A. Equality of Personal Characteristics

When two or more persons are said to be equal with respect to age or citizenship or race or income or aptitude or need, this simply means that they have the same age or nationality or color or income or ability or need¹—or that they are substantially similar in such respects. When Hobbes says that "nature has made men so equal in the faculties of the body and mind"² that anyone can kill, but not outwit, another, he means that all men have substantially the same physical and mental power, and that differences are insignificant. Persons of different age or race or ability are considered unequal in those respects. Human beings can be said to be equal or unequal only with respect to certain characteristics which must be specified. It is elliptic, and hence meaningless to say that "all men are equal." With respect to any given characteristic, some men may be equal, but all are unequal. The only characteristic which they all share is a common "human nature," but that is a tautological statement.

Equality and inequality of characteristics are no doubt descriptive concepts. Indeed, whether *A* and *B* have the same age or nationality or income can be empirically ascertained. So can assertions that *A* has greater ability or aptitude than *B*. These are characterizing value judgments:³ such statements are descriptive, not normative.

B. Equality of Treatment

Whether two or more persons are being "treated equally" or not is also an empirical question. *A* and *B* are treated equally by *C*, if *C* allots to *A* and *B* the same specified benefit (e.g., one vote) or burden (e.g., one year's military service), or the same amount of some specified benefit or burden (e.g., salary, tax burden). If *A* is let to vote but not *B*, if *A* is drafted but *B* exempted, if *A* receives a

higher salary than *B*, then *A* and *B* are treated unequally in those respects.

Whether *A* and *B* are to receive the same treatment will often depend on some general rule of distribution. *With respect to a given rule*, *A* and *B* are treated equally, not if they receive the same treatment, but if the rule is applied impartially to both. This is the concept of "Equality before the law, which lays down that we should treat each case in accordance with an antecedently promulgated rule."⁴ Equality before a law limiting suffrage to whites requires that any white and no black citizen be allowed to vote. Equality before the law does not demand that the law itself be egalitarian.

C. Egalitarian Rules of Distribution

Our concern is not with egalitarian or inegalitarian treatment relative to a given rule, but with the egalitarian or inegalitarian character of the rules themselves.

Like "just," "egalitarian" can be predicated only of those rules which stipulate how certain benefits or burdens are to be allocated among persons. One may ask whether it is morally right or wrong to legalize or to outlaw abortion or divorce, but not whether such policies are just or unjust,⁵ or whether they are egalitarian or inegalitarian. These latter categories can be applied to principles which stipulate how benefits (e.g., voting rights, salaries) or burdens (e.g. the duty to pay taxes, or to serve in the armed forces) are to be allotted.⁶

Rules of distribution have the general form: some specified benefit (e.g., franchise) or burden (e.g., a sales tax) is to be allocated or withheld from any person, depending on whether he has or lacks some specified characteristic (e.g., being a citizen over twenty-one, being white, buying cigarettes). Or: the amount of some specified benefit (e.g., salary) or burden (e.g., income tax) to anyone shall be a function of the amount or degree to which he has a certain characteristic (e.g., his ability, his income). Our question then is: Is there a criterion which permits us to classify any actual or conceivable

rule of distribution into egalitarian and inegalitarian ones, independently of any valuational or normative considerations?

II. TRADITIONAL CRITERIA OF EGALITARIANISM

Let us examine some of the criteria which have traditionally been applied, if often only implicitly.

A. Equal Shares to All

According to the most extreme view, a moral or legal system is egalitarian if *all* benefits and burdens are to be distributed in equal amounts to *all*. This is Aristotle's principle of numerical equality—"being treated equally or identically in the number and volume of things you get"⁷—applied to all things anyone is to receive—or has to relinquish. It is the utilitarian principle—"everybody to count for one, nobody to count for more than one"⁸—in the distribution of all benefits and burdens. Equal treatment of all in every respect has been advocated by some 19th century anarchists: equality of occupation (intellectuals to participate in manual work), of consumption (all to eat and dress alike), and especially of education would ultimately wipe out existing inequalities of personal characteristics such as those of talent and intelligence and would eventually mold a uniform human species.⁹

However, practically all rules of distribution are concerned with *certain* benefits or burdens, to be allocated to *certain* persons. Even principles as general as those of the American and French Revolutions proclaim that the same basic legal *rights* are to be given to all, and that means to all citizens in any given political system by their respective governments. If egalitarianism meant equal shares of everything to all, practically all existing rules would be inegalitarian.

B. Equal Shares to Equals¹⁰

Aristotle himself enlarged the criterion of egalitarianism to include rules which allot "equal

⁴ J. R. Lucas, *The Principles of Politics* (Oxford, Clarendon Press, 1966), p. 246.

⁵ Cf. William K. Frankena, "The Concept of Social Justice" in Richard B. Brandt (ed.), *Social Justice* (Englewood Cliffs, N. J., Prentice-Hall, 1962), p. 4.

⁶ I use the terms "benefits" and "burdens" to refer to anything which can be distributed to or exacted from several persons and which is generally valued by them either positively or negatively.

⁷ *Politics*, 1301 b.

⁸ John Stuart Mill, *Utilitarianism*, ch. V.

⁹ Cf. Isaiah Berlin, "Equality as an Ideal" reprinted in Frederick A. Olafson, *Justice and Social Policy* (Englewood Cliffs, N. J., Prentice-Hall, 1961), pp. 128-150; see p. 139.

¹⁰ To this point, and some of the following, cf. Felix E. Oppenheim, "The Concept of Equality," *International Encyclopedia of the Social Sciences*, vol. 5 (1968), pp. 102-107.

shares to equals"; i.e., equal shares of some specified kind to all who are equal with respect to some specified characteristic. Conversely, a rule is inequalitarian "when either equals are awarded unequal shares or unequals equal shares."¹¹

Here, the opposite criticism applies. Every existing, and even every conceivable rule of distribution turns out to be egalitarian in this sense, since every one allocates the same benefit or burden to all who have the same specified characteristic, but not to those who are unequal in that respect. Universal suffrage means that every adult citizen shall have one vote, but that minors and aliens shall have none. Suffrage to whites means that the right to vote is given to all white adult citizens, but not to colored persons. Conversely, an inequalitarian rule in this sense is a logical impossibility. A rule cannot stipulate that equals—in the sense of: those who have the characteristic specified by the rule—shall be awarded unequal shares, and unequals equal shares.¹² To practice racial discrimination is to give the same treatment to those of the same color, and to give unequal shares to those who are unequal with respect to this characteristic.

C. *Equal Shares to a Relatively Large Group*

Since every rule of distribution designates a certain class of persons who are to be treated equally, it could be argued, as it is by Isaiah Berlin, that one rule is more egalitarian than another if it insures "that a larger number of persons (or classes of persons) shall receive similar treatment in specified circumstances."¹³ To be more specific: a distribution of *benefits* is the more egalitarian, the larger the class of persons who receive it, as compared with the number of those excluded. Universal suffrage which excludes only minors and aliens is more egalitarian than a system which excludes Negroes in addition. Disenfranchising women is more inequalitarian than disenfranchising Negroes if the latter constitute less than half the population, but less inequalitarian if the majority is colored. Locke, who advocated equal political rights for property owners, was more egalitarian than his predecessors, but less so than later advocates of universal suffrage. On the other

hand, a rule which allots *burdens* is the more egalitarian the larger the class of persons on whom it is imposed. Exempting students from the draft is less egalitarian than drafting them also.

This criterion has the great advantage that egalitarianism and inequalitarianism become comparative concepts. From the point of view of empirical science, this is an advance from merely classificatory concepts, possibly leading to quantification. The disadvantage is that all rules of the type, "to each according to his need" would become highly inequalitarian, unless it so happens that a fairly large proportion of the population had the same, and high, degree of need. A sales tax would be very egalitarian; but a graduated income tax, very inequalitarian, since it divides taxpayers not merely into two classes but into a large number of brackets and imposes the greatest tax burden on the usually small number of those with the highest income. Only in the unlikely case that the great majority falls within the highest bracket would a graduated income tax become more egalitarian. Even the principle of equality of opportunity would, in spite of its name, be inequalitarian, since it provides greater advantages to those who lack certain opportunities than to those who already have them.

D. *Proportional Equality*

Yet, we are inclined to consider more benefits for the needier or a graduated income tax egalitarian. They would be, if egalitarianism were taken in the sense of Aristotle's "proportional equality" or "equality of ratios."¹⁴ A rule of distribution may be said to fulfill this requirement, provided the amount of benefit or burden allotted to anyone is a monotonically increasing function of the personal characteristic specified by the rule; i.e., the more of the characteristic, the greater the share. Any two persons are treated equally in this sense, provided the difference in the amount allotted to each is similarly correlated to the degree to which they differ with respect to the specified characteristic.

However, every conceivable rule would become egalitarian by this criterion, just as it would according to the principle of equal shares to

¹¹ *Ethics*, 1131 a.

¹² Cf. also John Rawls, "Justice as Fairness" reprinted in Olafson, *op. cit.*, pp. 80-107. "Now, that similar particular cases, as defined by a practice, should be treated similarly as they arise, . . . is involved in the very notion of an activity in accordance with rules" (p. 82).

¹³ Berlin, *loc. cit.*, p. 135.

¹⁴ *Politics*, 1301 b.

equals. Indeed, all rules of distribution not only allot "equal shares to equals" and "unequal shares to unequals," but also allots them "in proportion to" the latter's inequalities. Both rules, "to each according to his need" and "to each according to his height," assign different shares to different persons in the proportion in which they differ as to need or height. A flat rate and a graduated income tax both fulfill the requirement of proportional equality. Marx's ideal was the principle, "to each according to his need," rather than "to each according to his work." Yet, he did not deny that the latter rule, too, is egalitarian, since "the right of the producers [to receive means of consumption] is *proportional* to the labor they supply; the equality consists in the fact that measurement is made with an *equal standard*, labor."¹⁵ It is, therefore, an egalitarian principle, even though "it tacitly recognizes unequal individual endowment and thus productive capacity as natural privileges."¹⁶ Rules which establish only two categories are also egalitarian by this standard. Both universal suffrage and suffrage for whites only treat all persons in proportion to their inequality, with respect to the specified characteristic. Numerical equality is then but a special case of proportional equality.¹⁷

E. To Each According to His Desert

Aristotle sometimes contrasts equality, not with proportional equality in general, but with "equality proportionate to desert."¹⁸ Amounts of benefits are to be proportionate to the degree to which beneficiaries have—not whatever characteristic a rule might specify—but one specific characteristic; namely, relative desert. The more deserving a person, the greater his reward, and equal shares to persons of equal desert. Any criterion of distribution which disregards desert is then not truly egalitarian.

This time, it can, of course, not be argued that every rule turns out to be egalitarian. The criticism is rather that egalitarianism is here defined in valuational rather than in descriptive terms. Aristotle himself considers a distribution egalitarian in this sense, if "the relative values of the

things given correspond to those of the persons receiving."¹⁹ Now, the relative value of *things* given can usually be objectively ascertained and measured; and so can personal characteristics such as age or income, and even intelligence or aptitude for a certain task. On the other hand, the relative value of a *person* (receiving); i.e., the degree of his desert is clearly a matter of subjective valuation, not of objective assessment. Statements to the effect that *A* is more deserving than (or twice as deserving as) *B*, in the sense that *A* is of greater value or moral worth, are genuine, not characterizing, value judgments.

Implicit here is the Platonic-Aristotelian doctrine that men are essentially of unequal value or desert, in contrast to the later Stoic view of the equal worth or dignity of every human being. On the basis of the criterion under discussion, equality, e.g., of political rights, would be egalitarian to the latter and inequalitarian according to the former view. Again, if whites are considered "superior" to Negroes (in overall desert, not, e.g., in intelligence), then racial discrimination is egalitarian; the same policy would be inequalitarian to those who do not regard a person's worth as depending on his color.

F. Unequal Distributions Corresponding to Relevant Differences

At present the most widely held version of proportional equality is the following: a rule of distribution is egalitarian if, and only if, differences in allotments correspond to *relevant* differences in personal characteristics; in other words, provided the specified characteristic is relevant to the kind of benefits or burdens to be distributed. Thus, age and citizenship are said to be relevant to voting rights; it is, therefore, egalitarian to limit the franchise to adult citizens. Wealth is relevant to taxation; hence, a flat rate or a graduated income tax is egalitarian. Conversely, a rule is inequalitarian if it is either based on irrelevant differences of characteristics or disregards relevant ones. Sex or color or wealth are irrelevant to voting; restricting the franchise to men or whites or poll tax payers is

¹⁵ "Critique of the Gotha Program" in Lewis S. Feuer (ed.), *Karl Marx and Friedrich Engels, Basic Writings in Politics and Philosophy* (Garden City, N.Y., Anchor Books, 1959), p. 118. Italics in original.

¹⁶ *Ibid.*, p. 119.

¹⁷ Frankena considers rules satisfying the criterion of proportional equality *inegalitarian*. Cf. *Some Beliefs About Justice* (Department of Philosophy, University of Kansas, 1966), p. 7.

¹⁸ *Politics*, 1301 a.

¹⁹ *Politics*, 1280 a. The passage begins: "A just distribution is one in which. . ." However, "justice" and "equality" are synonymous to Aristotle. Cf. below.

inegalitarian. Wealth is relevant to taxation; hence, a sales tax is inegalitarian, since it taxes poor and wealthy buyers at the same rate.

Like personal desert, relevance of a personal characteristic is an evaluative, not a descriptive, term. While the ascription of characteristics such as a certain age or income to a person is a matter of fact, judgments to the effect that such characteristics are relevant or irrelevant to some kind of distribution are valuational, not factual. That age is relevant to voting, but color not, means nothing more than that it is just to require a minimum age for voting, but unjust to base franchise on color. It is inegalitarian—and that means that it is unjust—to treat persons unequally who share a “relevant” characteristic; but unequal awards to persons who differ in some “relevant” respect are egalitarian, i.e., just. Or, in a recent formulation, “a difference in treatment requires *justification* in terms of *relevant* and sufficient differences between the claimants.”²⁰ Advocates and opponents of racial discrimination are likely to disagree as to whether race is a “relevant” difference and whether discrimination is just. On the basis of the definition under discussion, they also would have to disagree as to whether such a policy is egalitarian.

This valuational interpretation of the concept of relevance has recently been challenged. For example, Bernard Williams holds it “quite certainly false” to claim “that the question whether a certain consideration is *relevant* to a moral issue is an evaluative question.” He argues as follows:

The principle that men should be differentially treated in respect to welfare merely on grounds of their color is not a special sort of moral principle, but (if anything) a purely arbitrary assertion of will, like of some Caligulan ruler who decided to execute everyone whose name contained three ‘R’s.’²¹

A racist’s advocacy of racial discrimination in welfare matters need not be arbitrary at all, but may well be rational—in the sense of consistent with his overall evaluations and other normative principles. To deny that such principles are *moral* ones is to apply the term “moral” itself in an

emotive sense to only those normative views to which one happens to subscribe.

Perhaps Williams means only to assert that the grounds on which such a normative principle would be defended, or criticized, reduces to purely empirical propositions. Indeed, he argues that, “if any reasons are given at all” for racial discrimination,

they will be reasons that seek to correlate the fact of blackness with certain other considerations which are at least candidates for relevance to the question of how a man should be treated: such as insensitivity, brute stupidity, ineducable irresponsibility, etc.²²

I do not deny that a statement such as “color is relevant to intelligence” is descriptive. It means that intelligence is a function of color, and this statement can be empirically tested, and disconfirmed. But here it is intelligence, not color, which is considered relevant; e.g., to voting rights. Unlike “color is relevant to intelligence,” “intelligence is relevant to voting rights” is normative, just as “color is relevant to voting rights” is normative.²³ It means that franchise ought to depend on intelligence or on color, and that a rule to that effect is just. To call a rule based on differences judged relevant *egalitarian* (rather than just) does not alter the normative character of the statement.

More recently, W. T. Blackstone has explicated the concept of relevance as follows:²⁴

To say “*x* is relevant,” when we are speaking about the treatment of persons, means “*x* is actually or potentially related in an instrumentally helpful or harmful way to the attainment of a given end and *consequently ought* to be taken into consideration in the decision to treat someone in a certain way.”

I agree with the author that the first part of this definition is descriptive and the second part prescriptive. But I disagree with “consequently.” I deny that a statement of the “is” type of relevance entails one of the “ought” type. Let us take his own example:

If, for example, race or color were cited as grounds for the differential treatment of persons in regard to

²⁰ Morris Ginsberg, *On Justice in Society* (Baltimore, Md., Penguin Books, 1965), p. 79. Italics added.

²¹ Bernard Williams, “The Idea of Equality” in Peter Laslett and W. G. Runciman (eds.), *Philosophy, Politics and Society* (Oxford, Basil Blackwell, 1962) [2nd series], pp. 110–131; see p. 113.

²² *Ibid.*

²³ Williams holds that “few can be found who will explain their practice merely by saying, ‘But they’re black: and it is my moral principle to treat black men differently from others.’” (*Ibid.*) Discrimination might be justified precisely in this way by segregationists who concede that Negroes are potentially no less sensitive, intelligent, and responsible than whites.

²⁴ W. T. Blackstone, “On the Meaning and Justification of the Equality Principle,” *Ethics*, vol. 77 (1967), pp. 239–253. The quotations are on p. 241. Italics added.

educational opportunities and it were shown that color or race has nothing to do with educability, then the factual presupposition of those who invoke these criteria would have been shown to be false and those criteria themselves to be irrelevant (in the factual sense of "relevant").

"Color is relevant to educability" is a factual statement, and "educability is relevant to educational opportunity" is normative. But the former statement does not entail the latter. Someone may agree that color "has nothing to do with"; i.e., is not relevant to, educability. Yet, he may hold without inconsistency that greater educational opportunities should be given to the more educable, or to whites, or that all should have the same education (i.e., that no group should receive preferential treatment). Blackstone himself concedes:

It could easily be the case that individuals agree on the factual part of a judgment of relevance (i.e., that certain facts are instrumentally related to certain goals) and yet disagree on the prescriptive part of that judgment (i.e., on what goal is desirable).

This seems to contradict the first-quoted statement ("consequently"). "Relevance" is not a descriptive criterion of egalitarianism as a characteristic of rules of distribution.

G. Unequal Distributions Which Are Just

Egalitarianism is sometimes defined directly in terms of justice (rather than indirectly, via relevance). According to a recent article by a political theorist, "the true opposite of equality is arbitrary, i.e., unjustifiable or inequitable inequality of treatment."²⁵ It would follow that justifiable or equitable inequality of treatment is "truly" egalitarian. Whether racial discrimination is egalitarian or inequalitarian would again depend on whether it is considered just or unjust.

This is an instance of what I should like to call "the definist fallacy in reverse." The definist fallacy itself consists of defining a value word; e.g., "good" or "desirable," by reference to descriptive terms; e.g., "happiness" or "approval." Now, if

"good" means the same as "conducive to happiness," or "desirable" the same as "approved by the majority," it would be self-contradictory to say that something which promotes happiness is bad, or that something is undesirable but approved by the majority. Aristotle's statement that "the unjust is unequal, the just is equal"²⁶ is another instance of this fallacy. Here the normative concept of justice is defined in terms of egalitarianism which Aristotle himself considers a descriptive term, as we have seen ("giving equal shares to equals"). Again, it is not self-contradictory to say that a graduated income tax is inequalitarian yet just.²⁷

Here we have the reverse procedure. Egalitarianism, a concept which we want to function descriptively, is defined by the normative concept of justice. If "rule *x* is egalitarian" means the same as "rule *x* is just (or justifiable or equitable)," then it is self-contradictory to consider a graduated income tax just and inequalitarian, or a sales tax inequitable but egalitarian.

Egalitarianism has been identified, even more broadly, with moral rightness. According to J. R. Lucas, a law may

be said to be unequal, in that the categories are wrongly specified, or the distinctions wrongly drawn, so that the law . . . discriminates between classes of people who ought not to be discriminated between.²⁸

Accordingly, people who disagree as to the rightness or wrongness of some discrimination are bound to disagree as to whether a law to that effect is egalitarian or inequalitarian.

H. Procedural Equality

Equality is also linked with justice by those who regard egalitarianism as a "procedural" principle: "Treat people equally unless and until there is a justification for treating them unequally."²⁹ Taken in this sense, "egalitarianism" does not refer to a characteristic of rules of distribution at all, but to a rule of distribution itself; namely: "All persons are to be treated alike, unless good reasons can be found for treating them differently."³⁰ It is true that this "Equality Injunction is not itself a positive

²⁵ W. Von Leyden, "On Justifying Inequality," *Political Studies*, vol. 11 (1963), pp. 56-70; see p. 67.

²⁶ *Ethics*, 1131 a.

²⁷ The same criticism seems to me to apply to Rawls's "Justice as Fairness," *loc. cit.* Rawls, too, defines justice in terms of equality (equal right to liberty, inequalities being justified only under certain conditions).

²⁸ *Op. cit.*, p. 256.

²⁹ Frankena, *loc. cit.*, p. 8.

³⁰ Monroe C. Beardsley, "Equality and Obedience to Law" in Sidney Hook (ed.), *Law and Philosophy* (New York, New York University Press, 1964), pp. 35-42; see p. 36.

rule of ethics, but a rule for adopting rules."³¹ It is, nevertheless, a *normative* rule (for adopting substantive rules).

This principle is not only purely normative but also purely procedural, compatible with whatever substantive discriminatory rules of distribution may be held "justified" or based on "good reasons." Such a criterion of egalitarianism does not enable us to classify substantive rules of distribution into egalitarian and inegalitarian ones.³²

The search for an adequate explication of the concept of equality has been fruitless so far. To repeat briefly: if egalitarianism were defined by "equal shares to all," hardly any rule would be egalitarian; if it meant "equal shares to equals" or "proportional equality," every rule would be; and any rule could be egalitarian on the basis of definitions referring to desert, or relevant differences, or justice. Procedural equality does not even designate a characteristic of rules of distribution. "Equal shares to a relatively large group" remains the least unsatisfactory definition, but I have indicated that its application leads to results which are often counter-intuitive. Indeed, even advocates of racial discrimination are likely to consider it inegalitarian (yet just) to restrict welfare benefits to whites regardless of need (even if the great majority of the population is white), but egalitarian (though unjust) to make welfare payments to the needy regardless of race (even if the needy are a small minority). I believe that it is possible to find a general descriptive criterion of egalitarianism which captures such distinctions.

III. PROPOSED CRITERION OF EGALITARIANISM

All the definitions we have examined so far consider only how much of some specified benefit or burden is to be allotted to any two persons, *A* and *B*. Rules of distribution may also be considered from the point of view of the end result. How much will *A* and *B* retain after the rule has been applied to them? How are benefits or burdens to be redistributed between *A* and *B*? We must then distinguish between three stages: (1) the original

distribution, (2) some rule of redistribution being applied, and (3) the final distribution resulting from 2. *Example 1*: (1) *A* has 8 units; *B* has 2; (2) take 3 from *A*; give 3 to *B*; (3) both *A* and *B* end up with 5.

A. Simple Criterion

A rule of redistribution might be said to be egalitarian, if it equalizes, or at least reduces the difference between initial holdings. *Example 1* would be an instance of an egalitarian redistribution, since the initial difference between the holdings of *A* and *B*, namely, 6 ($8-2$), is reduced to 0 ($5-5$). So would *Example 2*: Take 3 from *A* (who has 8) and nothing from *B* (who has 2)—since the difference between their holdings at the end ($5-2=3$) is smaller than it was at the start ($8-2=6$). Conversely, a redistribution which leaves previous inequalities of benefits or burdens unaffected or increases the difference would be inegalitarian. *Example 3*: Take 1 from *A* and 1 from *B* (the initial difference between their holdings, namely, 6, remains unaffected). *Example 4*: Take 1 from *A* and 2 from *B* (the difference increases from 6 to 7).

These examples show that a rule of redistribution can be said to be egalitarian or inegalitarian only relative to some previous distribution. Egalitarianism becomes an ordering concept, an advantage which it shares with the "least unsatisfactory definition" examined under IIC. With respect to a given distribution, a rule of redistribution is the more egalitarian, the smaller the difference between holdings at the end in comparison with those at the start. The redistribution in example 1 is more egalitarian than in example 2, more inegalitarian in example 4 than in example 3.

The examples also illustrate that equal allotments may lead to inegalitarian redistributions, and vice versa. A sales tax (example 3) is inegalitarian, since it weighs heavier on the poorer buyers and does not reduce differences in wealth. Conversely, a graduated income tax (example 2) tends to equalize previous holdings and is as such egalitarian by this criterion. This definition of an egalitarian rule does then remedy precisely the

³¹ *Ibid.* Similarly: "understood in this way, the principle of equality does not prescribe positively that all human beings be treated alike; it is a presumption against treating them differently, in any respect, until good grounds for distinction have been shown. . . . [It is] a rule of procedure for making decisions: Presume equality until there is reason to presume otherwise. But this is a formal, not a substantive rule." S. I. Benn and R. S. Peters, *Social Principles and the Democratic State* (London, George Allen & Unwin, 1959), p. 111. Yet it is a *rule*, not a characteristic of rules.

³² Furthermore, as we have seen, no rule stipulates that "all persons are to be treated alike." Every rule of distribution treats some people equally and others unequally.

defects of the definition examined under IIC.³³ Like the former, it is couched exclusively in descriptive terms, and is therefore valuationally neutral.

B. *More Adequate Criterion*

This rather simple criterion does, however, lead to counter-intuitive results in certain instances. *Example 5*: *A* has 97 units and *B* has 3; the difference between their holdings is 94. Taking 3 from *A* and 2 from *B* reduces their holdings to 94 and 1, respectively, and the difference between their holdings is now smaller than before; namely, 93 (instead of 94). Although more is taken from *A* who starts with more than from *B* who starts with less, we hardly would consider such a redistribution egalitarian.

Now, let us look at percentage differences between holdings. If the total of units at the beginning is 100, taking 3 from *A* (who has 97) and 2 from *B* (who has 3) reduces this total to 95. *A* is then left with about 99 per cent of this total (94/95), and *B* with about 1 per cent (1/95). The percentage difference between the final holdings of *A* and *B* (99—1) is 98; this is *larger* than 94, the percentage difference between their initial holdings. According to this criterion, the redistribution turns out to be *inegalitarian*.³⁴

This result is more in line with our general conception of egalitarianism. Indeed, if the difference between initial holdings is very large, taking more from those who have more does not necessarily make the redistribution egalitarian. I propose, therefore, to consider a rule of redistribution egalitarian if it reduces and inegalitarian if it increases the *percentage* difference between the holdings of those to whom the rule is being applied. With respect to a given initial distribution, a rule of redistribution is then the more egalitarian, the smaller the difference between the percentage

holdings at the end in comparison with the difference at the start. A sales tax is more clearly inegalitarian according to the present criterion than according to the previous.³⁵ Even a graduated income tax may be inegalitarian according to the present criterion (as in example 5). To be egalitarian, those in the highest brackets must pay proportionally very much, and those in the lowest very little (or nothing, as in example 2).³⁶

IV. SOME EGALITARIAN AND INEGALITARIAN PRINCIPLES

Let us examine a few of the more important rules of redistribution in the light of the proposed criterion of egalitarianism.

A. *Equalization of Wealth*

Full equalization of commodities, even when it is held desirable, is generally considered utopian. Even if this goal were realized at one moment, differences would soon reappear, if only because "men are unequal" as to personal endowments; hence, power and influence are bound to remain unequally distributed under every political and economic system. Equalization of wealth usually means merely reducing rather than removing existing inequalities of possessions. According to the proposed definition, this kind of redistribution, although less egalitarian, is egalitarian just the same. In Rousseau's words:

By equality, we should understand, not that the degree of power and riches be absolutely identical for everybody, but that . . . no citizen be wealthy enough to buy another, and none poor enough to be forced to sell himself.³⁷

On the other hand, "not even the equal distribution of money will lead to equal happiness."³⁸

³³ Definition IIC remains applicable when benefits and burdens are not quantifiable. E.g., extending the franchise from white to black citizens, or lowering the voting age, is egalitarian, since there is an increase in the proportion of citizens who receive the benefit relative to those who do not. This also satisfies the present criterion of egalitarianism, since the difference between initial position (having or lacking the franchise) is being reduced.

³⁴ I acknowledge my gratitude to T. J. Pempel, graduate student at Columbia University, for having suggested to me this improved criterion.

³⁵ The percentage difference between the initial holdings of *A* and *B* in example 3 is 60 (80—20). The total of their holdings is reduced from 100 per cent to 80 per cent. *A* ends up with 87.5 per cent (70/80), and *B* with 12.5 per cent (10/80). The percentage difference between their holdings has *increased* from 60 to 75 (whereas the absolute difference between their holdings remains the same).

³⁶ In example 2, the total of holdings is reduced to 70 per cent. *A* ends up with 50/70 = 70 per cent, and *B* with 20/70 = 30 per cent. The percentage difference is *reduced* from 60 to 50. The rule is egalitarian according to both criteria (whereas example 5 is egalitarian on the basis of the first and inegalitarian on the basis of the second criterion).

³⁷ Jean-Jacques Rousseau, *The Social Contract*, Bk. II, ch. XI.

³⁸ John Hospers, *Human Conduct* (New York, Harcourt, Brace & World, 1961), p. 424.

Besides, happiness or satisfaction or utility are not tangible benefits which can be distributed or redistributed to *A* and *B* by *C*, either equally or unequally.

B. Equality of Opportunity

Like utilities, opportunities cannot, strictly speaking, be given or distributed to *A* and *B* by *C*. "*A* has the opportunity to achieve *x*" means that there are no obstacles in his way of achieving *x*, so that he can do *x* if he wants to. *C* gives *A* the opportunity to reach *x* if he removes such obstacles and thereby enables *A* to achieve *x*, so that, whether *A* reaches *x* depends only on his native and acquired ability and on his effort. *A* and *B* have equal opportunity to win a race if they start from the same line. If *A* is initially behind *B*, he must be moved forward to the common starting line to have the same opportunity as *B*.

The principle of equality, or rather equalization, of opportunities is thus concerned with the redistribution of access to the various positions in society, not with the allocation of the positions themselves. The problem is: how to match individuals with unequal endowments with positions yielding unequal remuneration or power and prestige. The solution is to open them up to all on a competitive basis. The assumption is that, if everyone is given an equal start, the position everyone will occupy at the end will depend exclusively on how fast and how far he can run.

Classical liberalism held that equality of opportunity could be implemented by means of an equal allocation of the basic legal rights of "life, liberty, and property." If only legal privileges are abolished and equality of legal rights established, no obstacle will stand in the way of everyone's pursuit of happiness; i.e., everyone's ability to accede to the position commensurate with his highest ability.

Later it was realized that equality of rights is not sufficient to open up to the socially disadvantaged the opportunities open to the socially privileged. Unequal distributions are required to bring the former up to the common starting level: legal privileges and material benefits for the economically underprivileged, such as "head start" programs. To the extent to which such policies lead to an equalization of opportunities, they are egalitarian.

C. Equal Satisfaction of Basic Needs

The principle of equalization of opportunities is linked to another principle of equalization: the equal satisfaction of basic needs. While personal needs vary in kind and extent, there is a minimum of basic needs which are substantially identical for all in a given society at a given time. However, persons are unequal with respect to their *unsatisfied* basic needs. "Unequal distribution of resources would be required to equalize benefits in cases of unequal need."³⁹ The greater someone's unsatisfied basic need, the greater the benefits he receives. Those whose basic needs are already more nearly satisfied may not receive anything and may even have to give up some superfluities to provide for the former's necessities. The end result of such unequal distributions is, again, greater equalization of wealth and of opportunities.

D. To Each According to His Merit

Contemporary proponents of the democratic welfare state tend to combine the two egalitarian principles of equal satisfaction of basic needs and of equality of opportunity with another rule of redistribution: to each according to his merit. Once everyone's minimum needs have been taken care of, and all have been given an equal chance, the race is on, and the position everyone occupies at the end will depend only on his aptitude or "merit," again in theory at least. Unlike a person's "desert," his "merit" in the sense of proficiency at some specified task can in principle be objectively determined. But like "to each according to his desert," "to each according to his merit" is an inegalitarian rule of redistribution.

Schematically, we may then distinguish between the following stages: (1) an initial unequal distribution of commodities; (2) giving more to the needier, resulting in (3) a more egalitarian redistribution: equal satisfaction of basic needs, equality of opportunity; (4) from there on: an inegalitarian final redistribution: to each according to his merit.

This concept of equality is not only general and descriptive, but also valuationally neutral. For example, the author of *The Rise of Meritocracy* advocates "not an aristocracy of birth, not a plutocracy of wealth, but a true meritocracy of talent."⁴⁰ By the proposed criterion, all three of these principles are inegalitarian, the one he pro-

³⁹ Gregory Vlastos, "Justice and Equality" in Brandt, *op. cit.*, pp. 31-72; see p. 43.

⁴⁰ Michael Young, *The Rise of the Meritocracy: 1870-2033* (Baltimore, Md., Penguin Books, 1961), p. 21.

pounds as well as the two he rejects. On the other hand, advocates of "meritocracy" do in general not want to extend this principle to political participation; they remain in favor of equal suffrage, regardless of "merit."

This leads to the conclusion that modern democratic theory as a whole cannot be qualified as either egalitarian or inegalitarian, but is a mixture of both kinds of principles: equalization up to a certain level (by means of unequal distributions); inegalitarian redistributions beyond. It is, therefore, less inegalitarian than ideologies which base inequality of treatment on hereditary status, color, religion, or wealth.

There is, of course, no contradiction in calling meritocracy both inegalitarian and just. It may also be deemed unjust, yet desirable for other reasons—unjust because a person's merit depends in part on factors over which he has no control, such as innate intelligence and education or training (at least in the absence of full equality of educational opportunities)—desirable nevertheless on utilitarian grounds, because incentives to higher productivity will increase the welfare of all.

It has often been argued that men are equal and, therefore, egalitarianism just, or that inegalitarianism is equitable because men are unequal. For example, John Schaar, in a recent article, takes "the large discrepancy between the observed facts of inequality and the policy or value of equality as a serious intellectual embarrassment."⁴¹ As if it were inconsistent to hold that men should be given equal opportunities even though they are of

unequal intelligence—or unequal salaries in spite of their equal basic needs. Normative principles cannot be derived from factual generalizations; equality or inequality of some personal characteristic does not entail the desirability of either egalitarianism or inegalitarianism.

Mistaken arguments of this kind are often the result of confused language. There is the tendency to use factual statements for expressing normative views. We have seen that the allegation that "men are equal," if taken in the factual sense, is either meaningless, or tautological, or false. However, this adage serves more often as a rhetorical device to disguise the normative principle that men should be treated equally—in some respect which is often left unspecified. Then there is the temptation to use the factual statement that such and such a principle is egalitarian for the purpose of commending that particular rule. Conversely, valuational terms are being used to refer to some advocated goal; e.g., persons are to be treated according to their *desert*, or treated equally unless there are *relevant* differences, or unless unequal treatment is *justified*. When such value words are left unspecified, no substantive normative principle is being propounded.

Value words should be used exclusively to express the *advocacy* of some goal or principle; the *advocated* state of affairs should be characterized exclusively by descriptive terms. Following this practice would make for much-needed clarity in our moral discourse.

University of Massachusetts

Received November 19, 1968

⁴¹ John H. Schaar, "Some Ways of Thinking About Equality," *Journal of Politics*, vol. 26 (1964), pp. 867–895; see p. 868.

VI. RETURN OF THE LIAR: THREE-VALUED LOGIC AND THE CONCEPT OF TRUTH

BRIAN SKYRMS

I. INTRODUCTION

CHRYSIPPUS is reputed to have said that those who state the liar paradox "completely stray from word meanings; they only produce sounds, but they don't express anything."¹ The Chrysippian solution, by itself, however, will not do. If "is true"; "is false"; and "is meaningless" are related as global predicates which are incompatible and alternative the paradox persists, for the supposition that the liar sentence is meaningless leads to a contradiction just as inevitably as the suppositions that it is true or that it is false. For let the liar sentence be:

(a) *a* is not true

Then assume:

(1) "*a* is not true" is meaningless

But what is meaningless cannot be true, so then:

(2) "*a* is not true" is not true

Now *a* = "*a* is not true" and since an untruth by any other name is still an untruth, we have:

(3) *a* is not true

In all fairness to the Chrysippian approach, we cannot require that the concept of truth license the Tarski equivalences;

_____ if and only if "_____ " is true

for those equivalences lead in a straightforward way to bivalence, and the heart of the proposal under consideration is postulation of a third value which may be assumed by syntactically well-formed sentences composed of meaningful parts. But surely it is an irreducible part of the concept of truth that the corresponding arguments from

to
"_____ " is true

and conversely be *valid*. But even this weaker condition on the concept of truth (and it is weaker for validity of an argument may not guarantee logical truth of the corresponding conditional in three-valued logic) leads to trouble for from (3) we may now infer:

(4) "*a* is not true" is true

and since what is true cannot be meaningless we have

(5) "*a* is not true" is not meaningless

which contradicts assumption (1). Furthermore, "is false" may be replaced by "is such that its denial is true" and "is meaningless" may, *for these purposes*, be replaced by "is neither true nor false." Thus the full force of the liar paradox may be focused on the status of truth as a global predicate.

Let us say that '*T*' functions as a truth predicate (of sentences of *L*) in *L* if and only if *L* contains names of all sentences of *L* and the syntactical rules of *L* license the Tarski arguments for '*T*', and that '*T*' functions as a global truth predicate (of sentences of *L*) in *L* if and only if *L* contains names of all sentences of *L* and the syntactical rules of *L* guarantee that the result of substituting a name of any sentence of *L* for '*a*' in "*T*'*Ta*' ∨ *T*'~*Ta*' " is always a theorem of *L*. If *L* contains a global truth predicate (of sentences of *L*) and the standard logic holds in *L* and *L* has sufficient machinery to generate a liar sentence (i.e., if a sentence such as "*a* = '~*Ta*' " is a theorem of *L*) then *L* is inconsistent.

The lay of the land that I have just sketched has been known for some time² and periodically those unfamiliar with the literature have rediscovered

¹ Everet Wilem Beth, *The Foundations of Mathematics* (New York, Harper and Row, 1966), p. 25.

² See Rudolf Carnap, *The Logical Syntax of Language* (Paterson, N.J., Littlefield Admans, 1959), pp. 214-217; and Alfred Tarski, "The Concept of Truth in Formalized Languages" in *Logic, Semantics and Metamathematics* (Oxford, Clarendon Press, 1962), pp. 152-278.

fragments of the picture and announced the news amid great fanfare.³ Various attempts to disarm the liar by type-token or sentence-statement therapy have proved utter failures.⁴ The root difficulty clearly lies elsewhere. The basic philosophical concern is, then, to find a way of avoiding inconsistency which provides an adequate representation of the concept of truth.

II. THE CONCEPT OF TRUTH IN TARSKI HEIRARCHIES

No real progress was made toward the solution of the liar paradox until Tarski's work on the concept of truth in formalized languages. Tarski showed how to deal with truth while avoiding the liar paradox, essentially by fragmenting the global concept of truth into an infinite number of "local" truth-concepts in a language-metalanguage hierarchy. Each level only can refer to levels that are below it; each level can contain a truth predicate which covers those levels to which it can refer. The levels proceed upward *ad infinitum*. Thus "True" is replaced by "True in *L*"; "True in *ML*"; "True in *MML*"; etc. It is important to realize that this relativization of truth to a level of the hierarchy is different in kind from the relativization of truth to a language which is forced upon us by the possible ambiguity of symbol strings. Certainly, a certain sign design, *S*, may occur in two languages, *L*₁, *L*₂ with different meanings and thus might be true in *L*₁ and untrue in *L*₂. But this innocuous sort of relativization isn't what is at stake here. Put Platonistically, the relativization of truth to levels in a Tarski hierarchy is not simply a device to assure that no sentence express more than one proposition, but rather a restriction on what concepts and propositions can be expressed at all. In connection with this, note that it is a misconception to say that in a Tarski Hierarchy the liar paradox is explained away as a fallacy of equivocation. If it is argued that the Liar can't be formulated in *ML* because this involves an equivocation between

"true in *L*" and "true in *ML*," then it can be replied that *MML* contains a perfectly good truth predicate which covers both *L* and *ML*. The reason that the Liar can't be formulated in *ML* is that *that* truth predicate simply does not occur in *ML*. And the reason that it doesn't *isn't* that it would lead to a fallacy of equivocation but rather that it would lead to *inconsistency*. The fragmentation of truth in a Tarski hierarchy is real. We can't put Humpty Dumpty together again by "summing up" all the fragments for we can't talk about all levels of the hierarchy on any one level of the hierarchy, and if we attempt to do the summing up in some language, *L**, outside the hierarchy we are faced with the dilemma that either the resultant truth predicate won't cover *L** or else we resuscitate the liar within *L**. Thus, Tarski buys consistency at the price of giving up a global truth predicate.

This, however, is a heavy price to pay. We want to say that there is a common concept which underlies all Tarski's local truth predicates and to which they give but partial expression. Is this common concept then ineffable on purely logical grounds? Are we involved in nonsense when we attempt to utilize it? Or is even the belief in such a common concept merely an incoherent superstition? If so, then a great deal of our discourse, ordinary and philosophical, is reduced to pure rubbish. (Consider, for example, how to formulate the thesis that God is omniscient.) This is what makes the Chrysippian solution so attractive. Now, although the Chrysippian solution, *by itself*, will not do, a modified Chrysippian approach holds promise that we may be able to have a global truth predicate in a language which is consistent (and rich enough to be interesting).

III. THE MODIFIED CHRYSIPPIAN APPROACH

If we are to take the possibility of a Chrysippian solution seriously, we must allow for the possibility that sentences which appear to be true-or-false are not. (I shall use the word "neuter" as an abbrevi-

³ For instance, see Theodore Drange, "The Paradox of the Non-Communicator," *Philosophical Studies*, vol. 15 (1964), pp. 92-96, wherein the author makes the following extravagant claim: "My conclusion is that in the area of semantics contradictions are unavoidable. Once you avoid the contradiction associated with the liar, you cannot help but run into the contradiction associated with the non-communicator."

⁴ For instance, see Yehoshua Bar-Hillel, "New Light on the Liar," *Analysis*, vol. 18 (1957), pp. 1-6 and William W. Rozeboom, "Is Epimendies Still Lying?," *Analysis*, vol. 18 (1958), pp. 105-113. Saying that the liar sentence does not express a statement (or proposition) is simply a platonistic rephrasal of Chrysippus' solution. This assertion leads to a contradiction in exactly the same way as does the assertion that the liar sentence is meaningless. The type-token approach maintains that semantical predicates apply to tokens and calls into question the possibility of interpreting a language in which the liar is formulable such that all tokens of the same type have the same truth value. No plausible theory as to how the relevant tokens of the same type would vary their truth value has been developed. Later in this paper, I shall show that such desperate plays are unnecessary.

ation for "neither true nor false.") If we have no reliable way of deciding at the onset which sentences are neuter, we cannot simply prevent them from occurring within the system of logic. (This point takes on greater force if we discover, as we shall, that a sentence may be *contingently* neuter.) Thus, the underlying system of logic relevant to our considerations must be three-valued. (If anyone boggles at calling *neuter* a value, he may speak of a neuter sentence as lacking a value rather than having a third value. There is, as far as I am concerned, no substantive difference between these two modes of speech.)

Before we can proceed any further, we must decide how the truth value of a molecular sentence depends on the values of its constituent atomic sentences. Utilizing an idea of Bas van Fraassen,⁵ let us say that a molecular sentence is true (false) if and only if for any arbitrary assignment of truth and falsity to its neuter constituent atomic sentences, the molecular sentence would be true (false) on the customary truth table analysis. The sentential connectives are no longer value-functional on this approach, but the value of a molecular sentence is a function of the values of its constituent atomic sentences. Furthermore, those molecular sentences which are assigned *true* on every assignment of *true*, *false*, or *neuter* to the atomic sentences are just the classical propositional tautologies. I believe that van Fraassen's approach, for the first time, gives us a civilized way of doing three-valued logic.

The occurrence of a third value forces us to distinguish between three senses of validity which coalesce in two-valued logic:

- (A) Truth-preserving transformation
- (B) Non-falsity preserving transformation
- (C) Transformation that cannot lead directly from a truth to a falsehood

The fact that most of our experience in reasoning is concerned with non-neuter sentences could provide an explanation for an intuitive urge to take a form of argument which is valid in only one of these senses as valid in another.

Extreme care must be taken in evaluating supposed proofs, since a chain of transformations, each of which is valid in one of the foregoing senses may lead from a truth to a falsehood. *A fortiori*, *reductio ad absurdum* is not always legitimate.

For even if we derive a contradiction from an assumption by a chain of A-valid steps, all that we are entitled to conclude is that the assumption is not true. It does not follow that the assumption is false and its denial is true. The assumption and its denial may both be neuter.

* * *

Let us retain "is true" as a global truth predicate on the following sense:

- (1) The result of concatenating "is true" with a quotation-mark name of a sentence is itself either true or false.
- (2) The move from a sentence to the result of concatenating "is true" with the quotation of that sentence is A-valid (truth-preserving). Notice that it is not B-valid since it may take us from a neuter sentence to a false one.
- (3) The move from the result of concatenating "is true" with the quotation of a sentence to the sentence itself is A-valid. Note that it must also be B-valid, since by virtue of (1) the premiss must be either true or false.

* * *

Substitution of identities may take us from a non-self-referential context to a self-referential one. Since there are grounds for supposing that *certain* self-referential sentences may be neuter by virtue of their self-reference, we should allow that substitution of identities may take us from a true sentence to a neuter one. The converse substitution would take us from neuter sentence to a true one, and since the denial of a true sentence is false and the denial of a neuter sentence is neuter, substitution of identities may also take us from a neuter sentence to a false one. Thus, the best that can be safely assumed is that substitution of identities is C-valid. Note that so far we have not *assumed* that self-reference is always sinful, but only *allowed* for the possibility of an occasional sin on its part.

Attempts to dissolve the liar paradox by denying the truth of " $a = \sim Ta$ " (either by holding that it is false or that it is neuter) appear to be misguided. Consider the following familiar form of the liar paradox:

⁵ See Bas van Fraassen, "Singular Terms, Truth-Value Gaps, and Free Logic," *The Journal of Philosophy*, vol. 63 (1966), pp. 481-495. The reader will find that in my semantics for *S*, my notion of supervaluations and their use is somewhat different from van Fraassen's because of the possibility of an atomic sentence occurring within the scope of a quotation functor in a larger atomic sentence. The fundamental idea behind supervaluations, however, is the same.

The sentence in the box is not true.

The truth of

The sentence in the box = "The sentence in the box is not true."

is empirically ascertainable by exactly the same means we would use to ascertain the truth of

The sentence in the box = "Fang is a vicious dog."

if "Fang is a vicious dog" had been inscribed within the box, rather than "The sentence in the box is not true." (Note that *in this case* ' $\sim Ta$ ' is *contingently* self-referential. If, in accordance with Chrysippian intuitions, ' $\sim Ta$ ' turns out to be neuter by virtue of its self-reference, it will be *contingently* neuter. As I remarked above, it would then be absurd to try to exclude just those sentences which are neuter from our system of logic, for then whether a sentence was admissible would depend on whether it occurred within a box in such and such a place; whether someone used it as the first sentence of a certain article; whether someone uttered it at a particular time; etc.) Furthermore, there are cases where it is hard to deny that the relevant identity statement is necessarily true. The result of putting the quotation of "the result of putting the quotation of z for the last letter of the alphabet in z is not true" for the last letter of the alphabet in "the result of putting the quotation of z for the last letter of the alphabet in z is not true" is *identical with* "The result of putting the quotation of 'the result of putting the quotation of z for the last letter of the alphabet in z is not true' for the last letter of the alphabet in 'the result of putting the quotation of z for the last letter of the alphabet in z is not true' is not true."⁶ Since it does not seem that the identity statement itself is at fault, I shall assume that statements of identity are themselves either true or false.

⁶ Adapted from W. V. Quine, *Mathematical Logic* (New York, 1962), pp. 307-308. This version of the liar is also used by Pollock in "The Truth about Truth" (see footnote 9). In cases where the relevant identity statement is necessarily true, the liar paradox would work as well if a semantical predicate for logical necessity were substituted for the truth predicate. For although the analogue to the first Tarski argument (IIB) for necessity (i.e., from p you may infer $NQ(p)$) is not valid, in a sound system the weaker rule

If $\vdash p$ then $\vdash NQ(p)$

is valid. And if the relevant identity statement is demonstrable by logical means alone, we have an analogue to the liar stemming from a global concept of logical necessity. Such "modal liar" paradoxes have been investigated by Kent Wilson in his University of Pittsburgh doctoral dissertation. Also see R. Montague, "Syntactical Treatments of Modality, with Corollaries on Reflection Principles and Finite Axiomatizability," *Modal and Many-Valued Logics: Acta Philosophica Fennica*, fasc. 16, (1963), pp. 151-167.

Let us see how all this bears on the following sequence of sentences:

- | | |
|-------------------------------------|--------------------|
| 1. $a = "\sim Ta"$ | Premiss |
| 2. $T"\sim Ta"$ | Assume |
| 3. $\sim Ta$ | From 2, by IIC |
| 4. $\sim T"\sim Ta"$ | 1; 3 Sub. Identity |
| 5. $T"\sim Ta" \& \sim T"\sim Ta"$ | 2; 4 Conj. |
| 6. $\sim T"\sim Ta"$ | 2; 5 Reductio |
| 7. $T"\sim \sim Ta"$ | Assume |
| 8. $\sim \sim Ta$ | From 7, by IIC |
| 9. Ta | 8, D.N. |
| 10. $T"\sim Ta"$ | 1; 9 Sub. Identity |
| 11. $T"\sim Ta" \& \sim T"\sim Ta"$ | 6; 10 Conj. |
| 12. $\sim T"\sim \sim Ta"$ | 7; 11 Reductio |
| 13. $\sim Ta$ | 1; 6 Sub. Ident. |
| 14. $T"\sim Ta"$ | From 13 by IIB |
| 15. $T"\sim Ta" \& \sim T"\sim Ta"$ | 6; 14 Conj. |

If " $\sim Ta$ " is a liar sentence, then 1 is true. Suppose 2 is true. Then 3 must be true, since the move from 2 to 3 is A-valid. The move from 1 and 3 to 4 is only C-valid, so 4 may be either true or neuter. But since truth is a global predicate, and the denial of a true or false sentence must be true or false, 4 cannot be neuter by virtue of its form. Thus 4 must be true. If 2 and 4 are true, then their conjunction, 5, must also be true; but it is false by virtue of its form. This refutes the assumption that 2 is true. Thus 2 is either false or neuter. But it cannot be neuter by virtue of its form, so it must be false and its denial, 6, must be true. 6 is thus the result of a legitimate *reductio*. Similar reasoning establishes the legitimacy of the *reductio* that terminates on line 12. Line 6 says that " $\sim Ta$ " is not true; line 12 says that it is not false, so by conjoining them we have a proof that " $\sim Ta$ " is neuter. But doesn't this, itself, lead to a contradiction? Consider lines 13-15. 13 is gotten from 1 and 6 by a move which is only C-valid (i.e., a move which may take us from true sentences to a neuter one). And the move from 13 to 14, although A-valid, is not B-valid (it may take us from a neuter sentence to a false one). Thus the move from 1 and 6 to 14 is not valid in any sense, and the existence of a contradiction on line 15 is of no consequence.

But although the standard way of getting a contradiction out of the liar sentence has been blocked, can we be sure that there is no more devious way of deriving a contradiction under the modified Chrysippian approach? If such were the case, then the foregoing proof that the liar sentence is neuter would be no more compelling than a *reductio* proof that it is true in standard bivalent logic. The only way to lay such doubts to rest conclusively is to construct a formal system which embodies the basic ideas of the modified Chrysippian approach, reconstruct the argument at issue as a derivation within that system, and show that the system is consistent.

The System, S :

Vocabulary:

Individual Constants: a_1, a_2, \dots

Connectives: $\&, \vee, \sim, \supset, \equiv$

Predicates: $T, =$

Quotation Functor: Q

If a is an individual constant, then a is a *name*.

If a, β are names, then ' Ta '; ' $a = \beta$ ' are *atomic sentences*.

If a, β are sentences then ' $\sim(a)$ '; ' $(a \& \beta)$ '; ' $(a \vee \beta)$ '; ' $(a \supset \beta)$ '; ' $(a \equiv \beta)$ ' are *sentences*.

If a is a sentence, then ' $Q(a)$ ' is a *name*.

A *maximal occurrence of an atomic sentence* within a sentence is an occurrence of that atomic sentence which does not occur within a larger atomic sentence.

Axiom Schemata:

- (1) Schemata adequate to generate just the tautologies of the classical propositional calculus.
- (2) $TQ(p) \supset TQ(TQ(p))$
- (3) $\sim TQ(p) \supset TQ(\sim TQ(p))$
- (4) $TQ(p) \supset p$
- (5) $a = a$
- (6) $((TQ(a = \beta) \& TQ(p)) \supset \sim TQ(\sim p^*))$
where p^* is the result of substituting β for some or all occurrences of a in p and provided that a does not occur within the scope of a Q -functor in p .
- (7) $a = \beta \supset TQ(a = \beta)$
- (8) $\sim(a = \beta) \supset TQ(\sim(a = \beta))$
- (9) $(TQ(p) \& TQ(p \supset q)) \supset TQ(q)$

The foregoing yield axioms if names are substituted for ' a ' and ' β ' and sentences for ' p ' and ' q '.

Rules:

- (1) From $p; p \supset q$ to infer q
- (2) From p to infer $TQ(p)$

In order to get a system in which the liar sentence is embedded as strongly as possible the following axiom is added to S to form the system S_L :

$$a = Q(\sim Ta)$$

Although axiom schema (6) for substitution of identities is weaker than classical counterparts, certain sentences are given a special status in S which allows, for them, a more powerful use of schema (6). From (7) and (8) we get as theorems all sentences of the form

$$\sim TQ(\sim(a = \beta)) \supset TQ(a = \beta)$$

and likewise from (2) and (3) we get all sentences of the form:

$$\sim TQ(\sim TQ(p)) \supset TQ(TQ(TQ(p)))$$

Thus, from (5), (6), (7), and (8) we can get as theorems all sentences of the forms

$$a = \beta \supset \beta = a$$

and

$$(a = \beta \& \beta = \gamma) \supset a = \gamma$$

The special treatment accorded these "safe" sentences within S foreshadows a general method by which consistent two-valued theories can be embedded within three-valued logic with weak substitutivity of identity. The fundamental rationale of two-valued logic is to restrict the well-formed formulas of the system to a "safe" set of sentences which cannot include such singular sentences as the liar. The price paid for including the singular sentences into the set of well-formed formulas is the weakening of substitutivity of identity. It might be thought that paying this price renders the three-valued system too feeble to serve as a fundamental foundational logic. But if we can effectively identify a "safe" set of sentences to plug into a classical two-valued system, then we can treat those sentences with equal power in a three-valued system. All we need to do is to add the axioms of the two-valued system, together with axioms guaranteeing, in the manner indicated, that all the well-formed formulas of the two-valued system behave in a normal two-valued way. Thus, the gain in scope resulting from the three-valued approach entails no loss in power in dealing with those sentences which can be dealt with from a two-valued standpoint.

Schema (9) leads to a useful result concerning

substitution of derivable equivalents. From (9) we get as theorems all sentences of the form

$$TQ(p \supset q) \supset (TQ(p) \supset TQ(q))$$

and from here it is but a short step to the result:

$$\begin{aligned} \text{If } \vdash p \equiv q \text{ then } \vdash TQ(p) \equiv TQ(q) \\ \text{and } \vdash TQ(\sim p) \equiv TQ(\sim q) \end{aligned}$$

By induction, if O is any string of TQ s and negation signs:

$$\text{If } \vdash p \equiv q \text{ and } \vdash Op, \text{ then } \vdash Oq.$$

It also allows us to establish a quasi-deduction theorem for S . Instead of the classical deduction theorem

$$p \vdash q \text{ if and only if } \vdash p \supset q$$

we have:

$$p \vdash q \text{ if and only if } \vdash TQ(p) \supset TQ(q)$$

One half; If $\vdash TQ(p) \supset TQ(q)$ then $p \vdash q$, follows immediately from Rules (1) and (2) and schemata (4) and (9). The other half is gotten by induction on the length of a proof of q from p together with the axioms. Let $B_1 \dots B_n$ be such a proof. Then it is to be shown that $\vdash TQ(p) \supset TQ(B_i)$ for all $i \leq i \leq n$. Suppose $i = 1$. Either B_1 is an axiom or B_1 is p . If B_1 is an axiom, then $\vdash TQ(B_1)$ by rule (2) and $\vdash TQ(p) \supset TQ(B_1)$ by propositional logic [that is by use of instances of schemata (1) together with rule (1)]. If B_1 is p , then $TQ(p) \supset TQ(B_1)$ is an instance of schemata (1). But if $\vdash TQ(p) \supset TQ(B_k)$ for all $k < i$, then $\vdash TQ(p) \supset TQ(B_i)$. For either (1) B_i is an axiom or (2) B_i is p or (3) B_i follows by rule (1) from some B_j and B_m where $j < i$ and $m < i$ and B_m has the form $B_j \supset B_i$ or (4) B_i follows by rule (2) from some B_m where $m < i$ and B_i has the form $TQ(B_m)$. The first two cases have already been dealt with. In case 3 we have by hypothesis: $\vdash TQ(p) \supset TQ(B_j)$ and $\vdash TQ(p) \supset TQ(B_j \supset B_i)$. From schema (9) and propositional logic we have $\vdash TQ(B_j \supset B_i) \supset (TQ(B_j) \supset TQ(B_i))$. Thus by propositional logic we have $\vdash TQ(p) \supset (TQ(B_j) \supset TQ(B_i))$ and then $\vdash (TQ(p) \supset TQ(B_j)) \supset (TQ(p) \supset TQ(B_i))$ and then $\vdash TQ(p) \supset TQ(B_i)$. In case 4, we have by hypothesis: $\vdash TQ(p) \supset TQ(B_m)$. But $TQ(B_m) \supset TQ(B_i)$ will be an axiom since it will be an instance of schema (2). Thus, by propositional logic $\vdash TQ(p) \supset TQ(B_i)$. The quasi-deduction theorem is quite reasonable. S licenses an inference if and only if the statement saying that that inference is truth-preserving is provable in S . Counterexamples to the classical deduction theorem can be

located in inferences licensed by rule (2). If we had the classical deduction theorem we would have the Tarski equivalences and bivalence.

The liar sentence is provably neuter in S_L as follows:

- | | |
|---|--------------------------------|
| 1. $a = Q(\sim Ta)$ | Axiom |
| 2. $TQ(a = Q(\sim Ta))$ | 1, Rule (2) |
| 3. $(TQ(a = Q(\sim Ta)) \& TQ(\sim Ta)) \supset \sim TQ(\sim \sim TQ(\sim Ta))$ | Axiom: Schema (6) |
| 4. $TQ(\sim Ta) \supset \sim TQ(\sim \sim TQ(\sim Ta))$ | 2; 3, prop. logic. |
| 5. $\sim TQ(\sim \sim TQ(\sim Ta)) \supset \sim TQ(TQ(\sim Ta))$ | Sub. Equivalents, prop. logic. |
| 6. $TQ(\sim Ta) \supset \sim TQ(TQ(\sim Ta))$ | 4; 5, prop. logic |
| 7. $\sim TQ(\sim TQ(\sim Ta)) \supset TQ(TQ(\sim Ta))$ | Theorem of S (shown above) |
| 8. $\sim TQ(TQ(\sim Ta)) \supset TQ(\sim TQ(\sim Ta))$ | 7, prop. logic |
| 9. $TQ(\sim Ta) \supset TQ(\sim TQ(\sim Ta))$ | 6; 8, prop. logic |
| 10. $TQ(\sim TQ(\sim Ta)) \supset \sim TQ(\sim Ta)$ | Axiom: Schema (4) |
| 11. $TQ(\sim Ta) \supset \sim TQ(\sim Ta)$ | 9; 10, prop. logic |
| 12. $\sim TQ(\sim Ta)$ | 11, prop. logic |
| 13. $(TQ(a = Q(\sim Ta)) \& TQ(Ta)) \supset \sim TQ(\sim TQ(\sim Ta))$ | Axiom: Schema (6) |
| 14. $TQ(Ta) \supset \sim TQ(\sim TQ(\sim Ta))$ | 2; 13, prop. logic |
| 15. $TQ(Ta) \supset TQ(TQ(\sim Ta))$ | 7; 14, prop. logic |
| 16. $TQ(TQ(\sim Ta)) \supset TQ(\sim Ta)$ | Axiom: Schema (4) |
| 17. $TQ(Ta) \supset TQ(\sim Ta)$ | 15; 16, prop. logic |
| 18. $\sim TQ(\sim Ta) \supset \sim TQ(Ta)$ | 17 prop. logic |
| 19. $\sim TQ(Ta)$ | 12; 18, Rule (1) |
| 20. $\sim TQ(\sim \sim Ta)$ | 19, Sub. Equivalents |
| 21. $\sim TQ(\sim Ta) \& \sim TQ(\sim \sim Ta)$ | 12; 20, prop. logic |

Semantics for S : The consistency of S_L still needs to be established. This can be done in a purely syntactical way.⁷ But perhaps more insight can be attained by approaching the problem from a model-theoretic standpoint.

Let a *model*, M_i , for S consist of an ordered triple, $\langle A_i, B_i, f_i \rangle$, where A_i is a set of non-linguistic entities; B_i is the set of sentences of S ; and f_i is a denotation function which maps the names of S into $A_i \cup B_i$ such that the denotation of an expression consisting of a quotation functor followed by a sentence is that sentence (e.g., $f_i("Q(Ta)") = "Ta"$).

⁷ This has been done by Robert Meyer, by giving a mapping into S_5 . Schemata (3) and (8) are due to a suggestion of his.

Each sentence of S is assigned a *level relative to a given model* as follows:

If $f_i(a) \in A_i$ then $\lceil Ta \rceil$ is a sentence of the 0th level in M_i .

All identity sentences are sentences of the 0th level in M_i .

If $f_i(a)$ is a sentence of the N th level in M_i then $\lceil Ta \rceil$ is a sentence of the $(N+1)$ th level in M_i .

The level of a molecular sentence in M_i is the highest level attained by any of its constituent atomic sentences in M_i .

If a sentence is of no finite level in M_i it is of the level ω in M_i . ($\omega + 1 = \omega$)

Truth values of atomic sentences in a model are defined as follows:

$\lceil a = \beta \rceil$ is true in M_i iff $f_i(a) = f_i(\beta)$.

If $\lceil Ta \rceil$ is a sentence of the 0th level in M_i , then it is false in M_i .

If $\lceil Ta \rceil$ is of finite level and a is an individual constant; then if $f_i(a)$ is true in M_i , $\lceil Ta \rceil$ is true in M_i and otherwise $\lceil Ta \rceil$ is false in M_i .

If $\lceil Ta \rceil$ is of level ω and a is an individual constant, then $\lceil Ta \rceil$ is neuter in M_i .

If a is a quotation-functor name; then if $f_i(a)$ is true in M_i , $\lceil Ta \rceil$ is true in M_i and otherwise $\lceil Ta \rceil$ is false in M_i .

Truth values of molecular sentences in a model are assigned as follows:

A *Classical Valuation* in M_i for a molecule is an assignment of truth values to its constituent maximal occurrences of atomic sentences such that:

- (1) If an atomic sentence is true in M_i , it assigns truth to all maximal occurrences of that sentence within that molecule.
- (2) If an atomic sentence is false in M_i it assigns falsity to all maximal occurrences of that sentence in that molecule.
- (3) If an atomic sentence is neuter in M_i , it either assigns true to all maximal occurrences of that sentence in that molecule or false to all maximal occurrences of that sentence in that molecule.

A *molecule is true* in M_i if and only if that molecule is classically true (i.e., by a classical bivalent truth table analysis) under all classical valuations in M_i for that molecule.

A *molecule is false* in M_i if and only if that molecule is classically false under all classical valuations in M_i for that molecule.

A *molecule is neuter* in M_i if and only if it is neither true in M_i nor false in M_i .

The semantical analogue of the quasi-deduction theorem is the statement that $\lceil TQ(p) \supset TQ(q) \rceil$ is true in all models if and only if q is true in all models in which p is. That is:

$$\models \lceil TQ(p) \supset TQ(q) \rceil \text{ iff } p \models q$$

This result is forthcoming from our model theory. Going from left to right, suppose p is true in M_i . Then $\lceil TQ(p) \rceil$ is true in M_i . By hypothesis, $\lceil TQ(p) \supset TQ(q) \rceil$ is true in M_i . But $\lceil TQ(q) \rceil$ can only assume the values true and false. The assumption that it is false in M_i contradicts the hypothesis. Therefore it is true in M_i . But $\lceil TQ(q) \rceil$ can be true in M_i only if q is true in M_i . Going in the opposite direction, if p is true in M_i , then by hypothesis q is true in M_i . But then $\lceil TQ(p) \rceil$ and $\lceil TQ(q) \rceil$ are both true in M_i and so is $\lceil TQ(p) \supset TQ(q) \rceil$. If p is not true in M_i , then $\lceil TQ(p) \rceil$ is false in M_i and by the definition of truth values for molecular sentences, $\lceil TQ(p) \supset TQ(q) \rceil$ is true in M_i .

S can be shown to be *sound* on the basis of this model theory. The proof is trivial for the rules and all axioms other than those provided by schema (6). The axioms which fall under schema (6) can be shown to hold in all models as follows: All the relevant axioms may only assume the values true or false since their constituent maximal occurrences of atomic sentences must all be of the form of the truth predicate followed by a quotation-functor name and thus can themselves only assume the values true or false. Thus the only way one of the axioms can fail to be true in a model is if its antecedent is true and its conclusion false. This can only happen if $\lceil a = \beta \rceil$ is true; p is true; and p^* is false. Thus it suffices to show that for all M_i , if $\lceil a = \beta \rceil$ and p^* are both true in M_i , then p^* is either true or neuter in M_i . Now if a does not occur within the scope of a quotation functor in p , then it can only occur in maximal occurrences of atomic sentences of the form $\lceil a = \gamma \rceil$; $\lceil \gamma = a \rceil$; $\lceil Ta \rceil$. If $\lceil a = \beta \rceil$ is true in M_i , then substitution of β for a in maximal atomic occurrences of identity statements will leave the truth value of the identity statements undisturbed since the semantics for identity statements is classical. And since identity statements can only assume the values

true and false, such substitution of identity statements having the same truth value must leave the truth value of the molecule undisturbed. Thus, we have only to deal with maximal occurrences of atomic sentences of the form $\lceil Ta \rceil$. If $\lceil Ta \rceil$ and $\lceil T\beta \rceil$ are of finite level in M_i , then their truth values depend only on the truth values of $f_i(a)$ and $f_i(\beta)$ in M_i . But if $\lceil a = \beta \rceil$ is true in M_i then $f_i(a) = f_i(\beta)$. Thus $\lceil Ta \rceil$ and $\lceil T\beta \rceil$ must have the same truth value in M_i and since they are of finite level that value must be either truth or falsity. Substitution of one for the other cannot affect the truth value of the molecule. If $\lceil a = \beta \rceil$ is true, then $\lceil Ta \rceil$ and $\lceil T\beta \rceil$ must be of the same level, so the only remaining possibility is $\lceil Ta \rceil$; $\lceil T\beta \rceil$ both of the level ω in M_i . Four cases are possible: (1) a ; β both quotation mark names (2) a an individual constant; β a quotation mark name (3) a a quotation mark name; β an individual constant (4) a ; β both individual constants. In Case (1), if a ; β are both quotation mark names and $\lceil a = \beta \rceil$ is true, then $p = p^*$. In case (2) $\lceil Ta \rceil$ is neuter in M_i and $\lceil T\beta \rceil$ is either true or false in M_i . We can distinguish two subcases: (2-A) β replaces a in all occurrences of a in contexts of the form $\lceil Ta \rceil$ in p . (2-B) β replaces a in some of the occurrences of a in contexts of the form $\lceil Ta \rceil$ in p . In (2-A) the set of classical valuations in M_i for p^* must be a subset of the set of all classical valuations in M_i for p , and thus if p is true in M_i , p^* must be also. In (2-B) the sets of classical valuations in M_i for p and p^* must share at least one member, and thus if p is true p^* cannot be false. In case (3) $\lceil Ta \rceil$ is true or false in M_i and $\lceil T\beta \rceil$ is neuter. In both subcases of uniform substitution (3-A) and partial substitution (3-B) the best that can be said is that the set of classical valuations in M_i for p and p^* must share at least one member, and thus if p is true in M_i , p^* cannot be false in M_i . In case (4) $\lceil Ta \rceil$ and $\lceil T\beta \rceil$ are both neuter in M_i . In both cases of uniform substitution (4-A) and partial substitution (4-B) the set of classical valuations in M_i for p^* must be a subset of the set of classical valuations in M_i for p . Thus, if p is true in M_i , p^* must be also.

Since S is sound, and since there are plenty of models for S in which " $a = Q(\sim Ta)$ " is true, S_L is consistent. S is by no means complete, nor should it be. It should not be, because the model theory I have given for S is a conservative one; it makes many more sentences neuter in a model than need be. Some of this conservatism may be justified by the principle of sufficient reason. In a model in

which " $a = Q(Ta)$ " is true, my model theory says that " Ta " is neuter. But from the standpoint of consistency, it would be just as good to make it true or to make it false. (That is, if we add " $a = Q(Ta)$ " and " $TQ(Ta)$ " to S as axioms we get a consistent system. Likewise with " $a = Q(Ta)$ " and " $TQ(\sim Ta)$ ".) Another example is the two-step liar where " $a = Q(Tb)$ " and " $b = Q(\sim Ta)$ " are both true. Consistency requires only that we make at least one of the pair (" Ta ," " Tb ") neuter and either take the other at face value or make it neuter also. But the model theory gives the result that both are neuter. In these cases, I think that the model theory codifies sound intuitions. On the other hand, consider the case in which " $a = Q(\sim Ta)$ " and " $b = Q(\sim Ta)$ " are both true. My model theory makes not only " Ta " but also " Tb " neuter in this case. But here there is no reason why we cannot take " Tb " at face value without risk, in which case it is false. In such cases, it seems to me that the conservatism of the model theory is unfounded. The gap between S and the model theory I have given is, then, a symptom of defect on both sides. Some things should be provable in S that are not. The model theory makes some things logical truths and logical consequences which shouldn't be. Establishing the correct *rapprochement* is an open problem.

IV. CONCLUSION

It is now time to resolve an ambiguity in my initial characterization of a global truth predicate. Let us say that a truth predicate is *strongly* global if and only if every sentence of L has a name and it is provable in L that for *every* name of every sentence of L the resultant truth ascription statement is true or false; and that it is *mildly* global if and only if for every sentence there is *at least one* name such that resultant truth ascription statement is provably either true or false. Then what has been shown is that we can have a mildly global truth predicate in a consistent language in which the existence of the liar sentence is provable. (A strongly global truth predicate can, of course, be had in a consistent language too poor to be interesting. For instance, if the only names of sentences allowed were the sort of quotation names I use, then no liar sentence would be formulable and a strongly global truth predicate could be had cheaply.)

One might ask at this point what would be wrong with leaving substitutivity of identity alone

and letting the three-valued logic handle the liar all by itself.⁸ After all, the fact that we can derive a contradiction from each of the disjuncts of the tautology

$$TQ(\sim Ta) \vee TQ(\sim\sim Ta) \vee (\sim TQ(\sim Ta) \& \sim TQ(\sim\sim Ta))$$

simply shows that none of the disjuncts are true. Each of the disjuncts might be neuter, and yet the disjunction true. (This is the *only* other possibility given the handling of molecular sentences in three-valued logic here presupposed.) But note that this approach gives up even a mildly global truth predicate. And in so doing it poses a curious ineffability problem. We can say truly that the liar sentence is either true, false, or neuter. But if we try to say *which* it is, we can do no better than to utter a neuter sentence! One would have to be very very fond of classical substitutivity of identity to swallow this consequence cheerfully. But even this radical approach does not allow us to keep the classical *axioms* of substitutivity of identity. For if we let the axioms be of the form

$$(a = \beta \& p) \supset p^*$$

then we face the prospect of some of the axioms being neuter when p and p^* are neuter. And if we retreat to the form

$$(TQ(a = \beta) \& TQ(p)) \supset TQ(p^*)$$

we are faced with the same problem. Consider:

$$(TQ(a = Q(\sim Ta)) \& TQ(\sim Ta)) \supset TQ(\sim TQ(\sim Ta))$$

Where the first conjunct of the antecedent is true, the axiom must be neuter. For under this approach, the second conjunct of the antecedent is then neuter and the consequent must be either false or neuter. At best, substitutivity of identity might be salvaged as an inference rule. But here we would run up against another ineffability problem. For when we tried to say that certain arguments licensed by this rule were valid (that is, that if the premisses are true then the conclusion is true) we

might find that we could do no better than utter a neuter sentence.

I believe, then, that weakening of substitutivity of identity is part of the price that must be paid for a philosophically adequate solution to the liar paradox. This amounts to giving up Frege's principle that the denotation of a sentence (i.e., its truth value) is a function of the denotation of its constituents. For when " $\sim TQ(\sim Ta)$ " and " $\sim Ta$ " differ only in containing different names of the same thing, the first is true and the second neuter. Yet isn't this highly implausible? A rose by any other name would smell as sweet; if a thing has got a property then it has that property no matter what you call it; if saying one way that it has got that property is true, then saying another way that it has got that property is also true.

The foregoing contentions appear to me incontrovertible. But they merely show that we must also give up the principle that a well-formed sentence with meaningful constituents must be meaningful. If " $\sim Ta$ " said the same thing as " $TQ(\sim Ta)$ " then it would have to be true. But " $\sim Ta$ " *simply doesn't say anything at all*. The total failure of substitutivity of identity in referentially opaque contexts may show that we are talking about things other than what we think we are talking about. But the failure of classical substitutivity and success of weak substitutivity of identity (in what might be called referentially translucent contexts) merely shows that in certain cases where language turns about and bites its own tail, we fail to talk about anything at all.

But although some sentences fail to make the factual claim that they appear to make, for every factual claim there is some sentence which succeeds in making it. " $\sim TQ(\sim Ta)$ " succeeds in saying that the liar sentence is untrue. " $\sim Ta$ " appears to say the same thing, but in fact says nothing. The fact that the liar paradox is a paradox shows that some appearances in the area are deceiving. I submit that Chrysippus correctly identified the culprit.⁹

University of Illinois at Chicago Circle

Received December 10, 1968

⁸ This approach is taken by van Fraassen in "Presupposition, Implication and Self-Reference" (forthcoming) Sect. IV.

⁹ This paper grew out of comments on John Pollock's paper, "The Truth about Truth," delivered at the American Philosophical Association Western Division meetings in Chicago, May 1967. Earlier versions of this paper were read at the University of Western Ontario, The University of Pennsylvania, Indiana University, California State College at Fullerton, and the University of Illinois at Chicago Circle. Research was supported by the National Science Foundation.

VII. DISJUNCTIVE PREDICATES

DAVID H. SANFORD

I. THE PROBLEM

THE problem with which I am concerned may be introduced in several different ways. I shall begin by examining a passage from Panayot Butchvarov's *Resemblance and Identity: An Examination of the Problem of Universals*.

... A statement of the form "The universal of least generality instantiated in x and in y is of lower generality than the universal of least generality instantiated in w and in z " would seem to express faithfully the meaning of a statement of the form " x resembles y more than w resembles z ," assuming that both statements are about simple qualities.¹

For Butchvarov, two things instantiate the same universal if each has a specific quality which falls under the same genus. His treatment fails to rule out the following move: "If a thing has some specific quality F , and another thing has a specific quality H incompatible with F , then each thing has a specific quality which falls under the genus F or H . And if they have no specific quality in common, there is no universal of generality lower than F or H instantiated in both of them. Thus, on the above account, they do not resemble each other more, or less, than any other two things which have no specific quality in common and one of which has a specific quality incompatible with a specific quality of the other. Something yellow and something orange do not resemble each other with respect to hue more than something red and something green resemble each other."² This is clearly unacceptable. Butchvarov must find some way to rule out disjunctive qualities such as F or H as possible candidates for generic universals. Simple syntactical criteria are ineffective. It is useless to rule out predicates which have an explicit disjunctive form, for a predicate which has such a form can always be replaced by an equivalent one which does not. It is also useless to rule out predi-

cates equivalent to those which have an explicit disjunctive form, for a predicate which does not have such a form can always be replaced by an equivalent one which does. Butchvarov cannot appeal to a criterion which employs the notion of resemblance without begging the question. And such a criterion involves difficulties even for one who does not attempt to explicate resemblance with reference to the genus-species relation. It does no good to say that a predicate equivalent to one of the form " Q or R " is disjunctive if a thing which is Q need not resemble a thing which is R unless one can answer the charge that they resemble each other in both being either Q or R .

The problem of disjunctive predicates thus comes up when one attempts to characterize the relation between determinable and determinate, or between genus and species. Let us say that H putatively specifies G if everything H must be G . In his article, "On Determinables and the Notion of Resemblance,"³ John Searle suggests that G is nondisjunctive only if anything which putatively specifies G is logically related to everything else which putatively specifies G . (Two terms ' Q ' and ' R ' are said in this context to be logically related if and only if at least one of the following is satisfied: everything Q must be R ; everything R must be Q ; everything not Q must be R ; or nothing Q can be R .) Thus, "red or hard" fails the test for nondisjunctiveness. Although "red" and "hard" both putatively specify "red or hard," "red" and "hard" are logically unrelated.

Unfortunately, Searle's criterion rules out too much. Two terms are logically unrelated if their extensions satisfy all the following conditions: their intersection is not empty; they are not jointly exhaustive; and neither is a subset of the other. For example, "orange or yellow" and "yellow or green" are not logically related. But since both putatively specify "colored," "colored" fails Searle's test for nondisjunctiveness. Yet "colored"

¹ Bloomington and London, 1966, p. 129.

² This point is made somewhat more fully in my review of Butchvarov's book in *The Philosophical Review*, vol. 77 (1968), pp. 386-389.

³ *Proceedings of the Aristotelian Society*, Supplementary Volume XXXIII (1959), p. 148.

is Searle's paradigm of a nondisjunctive predicate.⁴

The problem of disjunctive predicates may also be approached via Nelson Goodman's New Riddle of Induction. All things to which predicate '*R*' applies are things to which any predicate equivalent to '*R* or *Q*' applies. This leads to certain embarrassments for theories of confirmation. Take t_0 to be some specific time in the future, say—to mark the 30th anniversary of Goodman's "A Query on Confirmation"—noon of July 4, 1976. "Grue" can be defined as follows:

x is grue at time $t = \text{DF}$ (x is green at t , and x was first examined before t_0) or (x is blue at t , and x was not first examined before t_0).

(Hereafter I shall omit the variable "at time t .") If all emeralds examined so far have been green, then all emeralds examined so far have been grue. We nevertheless take the hypothesis "All emeralds are green" as much better confirmed than "All emeralds are grue." Can we formulate a principle which accords with such a preference? It surely seems natural to say that "grue" is a disjunctive predicate and that "green" is not. The problem is to provide grounds for saying this which do not beg the question. It should be clear that predicates such as "grue" present puzzles which are not directly connected with theories of confirmation. No adequate characterization of the determinable-determinate relation will count "grue" as a determinate of "colored." And no one who is attracted by Butchvarov's explication of resemblance statements wants to say that something grue and first examined before t_0 and something grue and not first examined before t_0 resemble each other with respect to color just as much as any two green things resemble each other. Any solution relevant to Butchvarov's and Searle's problem is also relevant to Goodman's. I shall now present such a solution.

II. DISJOINT PREDICATES

Pakistan consists of two provinces—West and East Pakistan—approximately 1,000 miles apart, separated by the republic of India. The Dakotas consist of two states—North and South Dakota—right next to each other, not separated by anything. The area of Pakistan, unlike the area of the

Dakotas, is disjoint. An area is disjoint if it can be divided into two subareas the boundaries of which are completely distinct. Pakistan is disjoint because no part of the boundary of West Pakistan is part of the boundary of East Pakistan. But the single area called "the Dakotas" is not disjoint, for any two areas into which the Dakotas are divided will have some boundary in common. Part of the boundary of North Dakota, for example, is also part of the boundary of South Dakota.

I suppose that the boundaries both of Pakistan and of the Dakotas are quite sharply defined. But *pace* Frege, an area need not have a sharp boundary. The area occupied by members of the Dakota Indian Tribe, for example, need not. The west boundary of Dakota Indian Territory, for example, might be a certain mountain range; and there may be no relevant, nonarbitrary way of determining, for each point in the mountains, whether it is east of the range, west of the range, or precisely on the boundary between east and west. If we consider Dakota Indian Territory to consist of Santee Indian Territory and Teton Indian Territory, then Dakota Territory might be a disjoint area even though neither Santee Territory nor Teton Territory has a sharply defined boundary.

We are familiar with the representation of a predicate's extension by an area, as in Venn or Euler diagrams. If the predicate "red" is represented by a circle, red things are represented by points within the circle, and things which are not red are represented by points outside the circle. The concept of red is not absolutely sharp; the transition for red to orange is gradual; there is no point along the spectrum which separates red from not red. There are thus some *borderline cases*, things of which it would be wrong simply to say either "It is red" or "It is not red." Such borderline cases are naturally represented as points on the boundary of an area.

I should like to give a definition of a *disjoint predicate* which is analogous to the above definition of a disjoint area. A predicate is disjoint if it can be partitioned into two subpredicates which have no borderline cases in common. The predicate "red or green" is a disjoint predicate because it can be partitioned into the subpredicates "red" and "green," and no borderline case of "red" is also a borderline case of "green." I shall specify what I

⁴ This difficulty is also present in John Woods's "On Species and Determinates," *Nous*, vol. 1 (1967), pp. 243–254, which follows Searle's treatment closely in this and several other respects.

Another line of attack on the problem of disjunctive predicates is suggested in John Wisdom's *Problems of Mind and Matter* (Cambridge, 1963, second edition), pp. 29–31.

mean by "partitioning a predicate into two sub-predicates" with the intention of ruling out maneuvers which might be used to show that, on my definition, *every* predicate is disjoint.

First, the original predicate exhausts each of the subpredicates which in turn are jointly exhaustive of it and exclusive of each other. Thus, if '*P*' is the original predicate, and '*Q*' and '*R*' are the subpredicates into which it is partitioned:

$$(x)(Px \equiv (Qx \vee Rx)) \ \& \ (x)(Qx \supset \sim Rx)$$

Secondly, neither subpredicate can by itself exhaust the original predicate. Otherwise, we could, for example, partition the predicate "yellow" into subpredicates "yellow" and "round square." Since there are no borderline cases of round squares, there would be no borderline cases in common to the two subpredicates; and "yellow" would have to be counted as a disjoint predicate. So neither subpredicate may have a null extension:

$$(\exists x) Qx \ \& \ (\exists x) Rx$$

Predicates with non-null extensions but with no borderline cases, if there are any such, might still cause problems. Suppose that there is such a predicate '*S*' and that some things are both *P* and *S* while some other things are *P* but not *S*. Then perhaps '*P*' can be partitioned into "*P* and *S*" and "*P* but not *S*" which have no borderline cases in common. Let us write "*BPx*" for "*x* is a borderline case of '*P*'." The third restriction can then be written:

$$(x)(Qx \equiv (Px \ \& \ Sx)) \supset (\exists x) BSx$$

Since a subpredicate '*Q*' of '*P*' is always equivalent to "*P* & *Q*," it follows from this restriction that each subpredicate must have a borderline case:

$$(\exists x) BQx \ \& \ (\exists x) BRx$$

With our notation for borderline cases, the major requirement that the subpredicates not share borderline cases can be written:

$$(x)(BQx \supset \sim BRx)$$

This condition is not automatically satisfied by any partition. Despite its syntactical disjunctive form, the predicate "red or orange" is not shown to be disjoint by partitioning it into the subpredicates "red" and "orange"; for some borderline cases of "red" are also borderline cases of "orange."

As things stand, probably some predicates count as disjoint which we think ought not to count. Suppose that all ravens, and all creatures which

are borderline cases of "raven," are either definitely black, or, in the case of a few albinos, definitely white. The predicate "raven" could then be partitioned into the subpredicates "black raven" and "nonblack raven" which would have no borderline cases in common. It seems anomalous to call "raven" a disjoint predicate just because no ravens happen to be gray. We must distinguish what happens to be from what might be. Those who tolerate modalities are invited to prefix each universal quantifier by "It is necessary that" and each existential quantifier by "It is possible that" in all the formulae above and in all those to follow. Then, in addition to a set of extensional definitions, there will be as many sets of intensional definitions as there are relevant senses of "necessary" (or "possible"). Partitioning the predicate "raven" into "black raven" and "nonblack raven" would not show that it is a logically disjoint predicate, for it is logically possible that some borderline cases of "black raven" are also borderline cases of "non-black raven." Using borderline cases to characterize logical differences between predicates does not, of course, commit one to the view that such actual borderline cases are of any particular importance in our everyday use of the predicates.

III. DISCONNECTED PREDICATES

The area constituted by Colorado and Arizona is not disjoint, for the two states have one boundary point in common. The compound area is, however, disconnected. There are several ways of characterizing a disconnected area. A disconnected area can be partitioned into two subareas which, although they may have boundary points in common, have no boundary segments in common. Every point which is on the boundary of both subareas is also on the boundary of an area which is part of neither subarea. In other words, every point on the boundary of either subarea is also on the boundary of the compound area. For example, every point on the boundary of Colorado, and every point on the boundary of Arizona, is also on the boundary of the compound, disconnected area constituted by the two states. The Dakotas, in contrast, do not constitute a disconnected area. The boundary between North and South Dakota is not part of the boundary of the area formed by taking the two states together.

A *disconnected predicate* is one which can be partitioned into two subpredicates such that every borderline case of each subpredicate is also a

borderline case of the original predicate. This condition,

$$(x)((BQx \vee BRx) \supset BPx)$$

replaces the stronger condition of a disjoint predicate that the two subpredicates have no borderline cases in common. Unless we add the condition, as I see no reason to do, that the subpredicates of a disconnected predicate must share some borderline cases, all disjoint predicates are disconnected. Because the subpredicates which result from the partition of a predicate are exclusive of each other, both disjoint predicates and predicates which are disconnected but not disjoint may be said to be *exclusively disjunctive predicates*.

Is "grue" an exclusively disjunctive predicate? Its definition mentions subpredicates into which it can be partitioned. Since something can be a borderline case of "green," of "blue," and of "was first examined before t_0 ," it is not a disjoint predicate. Is it a disconnected predicate? Are all the borderline cases of "green and first examined before t_0 " as well as all the borderline cases of "blue and not first examined before t_0 " also borderline cases of "grue"? I think it is significant that our definition of "grue" is insufficient to settle this question.

A natural first reaction to the question why "green" rather than "grue" should be projected under normal circumstances is that the second is complex in a way that the first is not. "Grue" is, after all, defined in terms of two color predicates and a temporal predicate. Goodman's response to this move is so well known that his puzzle is often called the "grue-bleen problem." The definition of "bleen" is exactly like that of "grue" except that the terms "blue" and "green" are interchanged:

x is bleen = DF (x is blue, and x was first examined before t_0) or (x is green, and x was not first examined before t_0)

Granted that "grue" is defined with reference to "blue" and "green," we may now show, so the response goes, that "green" may as easily be defined with reference to "grue" and "bleen":

x is green = DF (x is grue, and x was first examined before t_0) or (x is bleen, and x was not first examined before t_0)

Let us substitute extensionally equivalent predicates in the above disjunction:

x is green = DF (x is green, and x was first examined before t_0) or (x is green, and x was not first examined before t_0)

No definition of "disjunctive predicate" on which "grue" is disjunctive will be of much interest if "green" also counts as disjunctive. The above partition does not show "green" to be disjoint. Something which is green but a borderline case of "first examined before t_0 " is a borderline case of both subpredicates into which "green" is partitioned. And it seems that the above partition also does not show that "green" is disconnected. For the borderline case just mentioned is green—and thus not a borderline case of "green"—so some borderline cases of the subpredicates are not also borderline cases of the original predicate. But this shows that something is wrong with the so-called "definition" of "green." If we look just at the definition, it seems that something which is a borderline case of each subpredicate ought also to be a borderline case of their disjunction. If something is neither definitely ' Q ' nor definitely ' R ', how can it be definitely " Q or R "?

Consider the formula:

$$(x)(Px \equiv ((Px \& Sx) \vee (Px \& \sim Sx)))$$

Our definition of "green" in terms of "green and first examined before t_0 " and "green and not first examined before t_0 " is of this form. And since the formula is a tautology, what could be wrong with the definition? The trouble is that the formula should not be regarded as a tautology in this context. Logic quite properly usually ignores borderline cases. But we are not ignoring them; and there is no reason, in general, why something which is P may not be a borderline case of ' S '. Previously it appeared that any predicate ' P ' could be partitioned into subpredicates " P and S " and " P and not- S ." But this will not do if we choose not to ignore borderline cases. And although the following formula is not incorrect

$$(x)(Px \equiv ((Px \& Sx) \vee (Px \& \sim Sx) \vee (Px \& BSx)))$$

it is of little help in dealing with such predicates as "grue." Suppose we give the following "definition" of "green":

x is green = DF (x is green, and x is first examined before t_0) or (x is green, and x is not first examined before t_0) or (x is green, and x is a borderline case of "first examined before t_0 ")

Two questions arise. (1) How are we to replace occurrences of "green" with occurrences of "grue" or "bleen" in the *definiens*? We know how to handle the first and second disjuncts, of course, but we are presently unable to deal with the third. And

similarly, (2) how can we extend our definition of "grue"? A third disjunct " x is grue and x is a borderline case of 'first examined before t_0 '" is unhelpful, and not just because it makes the definition circular. Our problem is that so far we have no idea what it is to be both grue and a borderline case of "first examined before t_0 ." Until we can stipulate this, there will be an asymmetry between "grue" and "green"; for there is no particular problem in a thing's being both green and a borderline case of "first examined before t_0 ."

Let us return to our spatial analogy. Something is in the Dakotas if and only if it is in North Dakota, or is in South Dakota, or, though it is both on the boundary of North Dakota and on the boundary of South Dakota, is not on a boundary of an area which includes neither part of North Dakota nor part of South Dakota. If something is on the boundary of Minnesota, for example, then even if it is on the boundaries of both North and South Dakota, it is not in the Dakotas, but is only on the boundary of the Dakotas. We may replace the earlier, simpler biconditional given in the specification of a partition with:

$$(x)(Px \equiv (Qx \vee Rx \vee (BQx \& BRx \& \sim B(\sim Q \& \sim R)x)))$$

The other conditions listed in the specification of a partition remain the same. ' Q ' and ' R ' are still to be called "subpredicates." When ' P ' is shown to be disjoint by a partition, then nothing satisfies the third disjunct. It is important to see that not all partitions show that a predicate is exclusively disjunctive. "Green," for example, may be partitioned into the subpredicates "green and first examined before t_0 " and "green and not first examined before t_0 ." We have already seen that this partition does not show "green" to be disjoint; and it also does not show "green" to be disconnected, since some borderline cases of the subpredicates are not borderline cases of "green." Partitions of the above form have the advantage that it no longer looks as if all the borderline cases of a subpredicate ought to be borderline cases of the original predicate.

We can now solve the problem of formulating a definition of "grue" which does not ignore the possibility of borderline cases. If our original definition is taken as adequate, then "grue" is disconnected. All the borderline cases of "green and first examined before t_0 " as well as all the borderline cases of "blue and not first examined before t_0 " are

borderline cases of "grue." We have the means, however, to define "grue" so that something can both be grue and a borderline case of "first examined before t_0 ." We could cast the definition in the same form as the above biconditional for a partition, but this is unnecessarily complicated. We may more simply add to the previous definition that something which is a borderline case both of "green and first examined before t_0 " and of "blue and not first examined before t_0 " is grue unless it is also a borderline case of "neither green nor blue." "Grue" is then not a disconnected predicate. If "bleen" is similarly defined so as not to be disconnected, then the symmetry of interdefinability between "grue" and "bleen" on one hand and "green" and "blue" on the other is restored. The definition of "green" may be cast in the same form as the above biconditional for a partition, with ' P ' interpreted as "green," ' Q ' as "grue and first examined before t_0 ," and ' R ' as "bleen and not first examined before t_0 ." The complicated third disjunct of this definition is satisfied only by things which are green and borderline cases of "first examined before t_0 ."

Even if "grue" and "bleen" are defined so as not to be disconnected, there is still a respect in which they are disjunctive and "green" and "blue" are not. This respect will be discussed in Sect. V.

IV. INCLUSIVELY DISJUNCTIVE PREDICATES

Searle was bothered by "red or hard." So far nothing has been done to characterize a respect in which such predicates are disjunctive. "Red or hard" cannot be partitioned into the subpredicates "red" and "hard"; for the subpredicates which result from a partition are exclusive, and "red" and "hard" are not exclusive.

I shall approach this problem by considering a predicate which I want not to consider inclusively disjunctive even though it is equivalent to the disjunction of two overlapping predicates. Although I see no logical reason why two determinate predicates of the same determinable—or two species predicates of the same genus—might not overlap (and thus I cannot accept Searle's solution to the problem of disjunctive predicates), I must admit that natural examples of such overlapping predicates are hard to find. My purpose can be served by an artificial example. Let us understand "redange" to apply to red things, orange things, and things which are borderline cases both of

"red" and of "orange" but not of "neither red nor orange." And let us similarly understand "yellange" to apply to yellow things, orange things, and things which are borderline cases both of "yellow" and of "orange" but not of "neither yellow nor orange." Now consider the predicate "redange or yellange." It applies to colors in the spectrum between red and yellow inclusive. Although it is broader than most of our color terms, there is intuitively no reason to call it a disjunctive predicate. Something which is a borderline case both of "orange" and of "brown" is a borderline case both of "redange" and of "yellange." It is also a borderline case of the compound predicate "redange and yellange."⁵ But it is not a borderline case of either "redange and not yellange" or "yellange and not redange." I take this rather complicated point to be important and shall try to clarify it by a spatial analogy.

The lines across a football field are numbered from both ends toward the middle: 10, 20, 30, 40, 50, 40, etc. Let us regard a football field as composed of two intersecting areas, *A* and *B*. *A* is the area from one goal line to the further 40-yard line, and *B* is the area from the other goal line to the other 40-yard line. Thus the area between the two 40-yard lines belongs both to *A* and to *B*; the area between one goal line and the 40-yard line nearer to it belongs to *A* but not to *B*; and the area between the other goal line and the 40-yard line nearer to it belongs to *B* but not to *A*. Now consider the intersection of the 50-yard line with one of the sidelines. This point is on the boundary both of *A* and of *B*. But it is not on the boundary of the area which belongs to *A* but not to *B*, nor is it on the boundary of the area which belongs to *B* but not to *A*. If something were to move just a short distance from this point so as to be within *A*, it would also be within *B*, and *vice versa*. Areas *A* and *B* share a segment of boundary—the sideline between the two 40-yard lines—which is also a boundary of the football field. The fact that the two areas share such a segment of boundary makes it natural to consider one as an extension of the other. Analogously, if something is a borderline case both of "redange" and "yellange," what kind

of borderline case it is of one is determined by what kind of borderline case it is of the other.

Compare "redange or yellange" with "red or hard." Some things are red and hard; some are red and not hard; and some are hard and not red. But if something is a borderline case both of "red" and of "hard," what kind of borderline case it is of one is irrelevant to what kind of borderline case it is of the other. It is a borderline case both of "red and not hard" and of "hard and not red." This is the peculiarity I shall exploit in defining inclusively disjunctive predicates. The spatial analogue of such a predicate is an area composed of two intersecting areas whose boundaries have no segments in common, although they may have any number of points in common. A two-circle Venn diagram, for example, has two points which are common to the boundaries of both circles; and each point is also on the boundaries of both crescent-shaped areas which contain the points which are within one circle but not within the other.

A predicate is *split* into subpredicates if it exhausts each of the subpredicates which in turn are jointly exhaustive of it and not exclusive of each other. (I use "split" rather than "partitioned" here simply because I have used "partitioned" already.) Thus, if '*P*' is the original predicate, and '*Q*' and '*R*' are the subpredicates into which it is split:

$$(x)(Px \equiv (Qx \vee Rx)) \ \& \ (\exists x)(Qx \ \& \ Rx)$$

A predicate is *inclusively disjunctive* if it can be split into two subpredicates which, although they have borderline cases in common,

$$(\exists x)(BQx \ \& \ BRx)$$

what kind of borderline case something is of one subpredicate is irrelevant to what kind of borderline case it is of the other:

$$(x)((BQx \ \& \ BRx \ \& \ B(Q \ \& \ R)x) \supset \\ (B(Q \ \& \ \sim R)x \ \& \ B(R \ \& \ \sim Q)x))$$

"Red or hard" is inclusively disjunctive on this definition, but, so far as I can see, "redange or yellange" is not.

⁵ Not all inferences of the form

$$BP a \ \& \ BQ a \therefore B(P \ \& \ Q) a$$

are valid. For every borderline case of '*P*' is a borderline case of "not-*P*," and inferences of the form

$$BP a \ \& \ B\sim P a \therefore B(P \ \& \ \sim P) a$$

are clearly invalid. There are no borderline cases of self-contradictory predicates. Some inferences of the form

$$B(P \ \& \ Q) a \therefore BP a \ \& \ BQ a$$

are also invalid. If something is a borderline case of '*P*' and is definitely *Q* (and thus not a borderline case of '*Q*'), then it is a borderline case of "*P* & *Q*."

V. SKEW PREDICATES

Let us contrast the Dakotas with the area constituted by South Dakota and Wyoming together. South Dakota shares a segment of boundary both with North Dakota and with Wyoming. Yet there is a difference. The boundary between South Dakota and Wyoming is longitude $104^{\circ} 3'$ West. Something can be at this longitude and be on the boundary of Wyoming without being on the boundary of South Dakota. And something can be on the boundary of South Dakota and be at this longitude without being on the boundary of Wyoming. Wyoming and South Dakota together constitute what I call a skew area. The Dakotas, North and South Dakota together, in contrast, do not constitute a skew area. If something is at latitude $45^{\circ} 57'$ North, it is on the boundary of North Dakota if and only if it is on the boundary of South Dakota.

An analogous definition of a *skew predicate* can be provided. Suppose that a predicate '*P*' is partitioned into subpredicates '*Q*' and '*R*.' Suppose also that the incompatibility of the subpredicates can be revealed by biconditionals of the following form:

$$(x)(Qx \equiv (Tx \& Sx)) \& (x)(Rx \equiv (Ux \& \sim Sx))$$

Call '*T*' and '*U*' "sub-subpredicates," and call '*S*' the "boundary predicate." Suppose further the predicates formed by disjoining a sub-subpredicate with the boundary predicate are inclusively disjunctive. In other words, both "*T* or *S*" and "*U* or *S*" are inclusively disjunctive predicates. The crucial condition follows:

$$\begin{aligned} &(\exists x)(BQx \& BSx \& \sim BRx) \vee \\ &(\exists x)(BRx \& BSx \& \sim BQx) \end{aligned}$$

If something is a borderline case of one subpredicate and also a borderline case of the boundary predicate, it does not follow that it is a borderline case of the other subpredicate. If a predicate such as '*P*' satisfies all the above conditions, it is a *skew predicate*. Let us say that skew predicates, as well as those which are disjoint or disconnected, are exclusively disjunctive.

"Grue" is a skew predicate. Subpredicates into which it is partitioned, "green and first examined before t_0 " and "blue and not first examined before t_0 ," are already of the required form. The predicates "green or first examined before t_0 " and "blue or first examined before t_0 " are both inclusively

disjunctive. And the crucial condition is satisfied. If something is a borderline case both of "green and first examined before t_0 " and of "first examined before t_0 ," it does not follow that it is a borderline case of "blue and not first examined before t_0 ." Also, if something is a borderline case both of "blue and not first examined before t_0 " and of "first examined before t_0 ," it does not follow that it is a borderline case of "green and first examined before t_0 ." Both disjuncts of the crucial condition are satisfied here, although only one need be for the original predicate to be skew.

A similar treatment fails to show that "green" is a skew predicate. When "green" is partitioned into "grue and first examined before t_0 " and "bleen and not first examined before t_0 ," the crucial condition is not satisfied. A borderline case of "first examined before t_0 " is a borderline case of "grue and first examined before t_0 " if and only if it is a borderline case of "bleen and not first examined before t_0 ." This, of course, is not a conclusive proof that "green" is not a skew predicate. The possibility remains that some other partition of "green" satisfies the crucial condition. I both hope and believe that this is not the case. The requirement that the predicate formed by disjoining a sub-subpredicate with the boundary predicate be inclusively disjunctive rules out all the maneuvers I could think of which might show that any predicate is skew. I have formulated several sufficient conditions for a predicate's being disjunctive, and in each case I have attempted to anticipate and thwart ingenious machinations designed to show that all predicates satisfy these conditions. Although I cannot prove that any predicate is not disjunctive, it seems reasonable to regard a predicate as non-disjunctive if there is no apparent way of proving the opposite.

Predicates of a sort I have not considered might convince us that further sets of sufficient conditions for disjunctiveness should be formulated. For example, perhaps Goodman could replace "grue" with a similar predicate whose subpredicates exclude each other but do not satisfy my requirements for a partition because at least one of them cannot have borderline cases. I do not see how such a predicate could be introduced or, if all empirical predicates can have borderline cases, how it could be relevant to the New Riddle of Induction.⁶ At least one of the subpredicates must be empirical if, as is necessary for a possible inductive inference to be in question, we can have empirical evidence that

⁶ See Max Black, "Reasoning with Loose Concepts," *Dialogue*, vol. 2 (1963), pp. 1-12.

the predicate does or does not apply to certain things. And, since the subpredicates exclude each other, the other subpredicate must also be empirical. But even if the contemplated maneuver is possible in a way I have not anticipated, supplementary principles such as the following could be invoked to escape the difficulty: If ' P ' is shown to be an exclusively disjunctive predicate by a partition into subpredicates ' Q ' and ' R ', and if $(x)(Sx \supset Qx)$ and $(x)(Tx \supset Rx)$, then any predicate equivalent to " S or T " is exclusively disjunctive. If ' P ' is shown to be an inclusively disjunctive predicate by a split into subpredicates ' Q ' and ' R ', and if $(x)(Qx \supset Sx)$ and $(x)(Rx \supset Tx)$, and if ' S ' and ' T ' are logically independent, then any predicate equivalent to " S or T " is inclusively disjunctive. Both these principles allow a predicate of the form " S or T " to be disjunctive even though neither ' S ' nor ' T ' has any borderline cases.

VI. THE NEW RIDDLE OF INDUCTION

Since there seems to be no general agreement about what form an adequate solution of Goodman's puzzle should take, I shall try to follow the line that Goodman himself recommends.

A rule is amended if it yields an inference we are unwilling to accept; an inference is rejected if it violates a rule we are unwilling to amend. The process of justification is the delicate one of making mutual adjustments between rules and accepted inferences; and in the agreement achieved lies the only justification needed for either.

All this applies equally well to induction. An inductive inference, too, is justified by conformity to general rules, and a general rule by conformity to accepted inductive inferences. Predictions are justified if they conform to valid canons of induction; and the canons are valid if they accurately codify accepted inductive practice.⁷

Goodman points out that certain theories of confirmation permit inferences we are unwilling to accept. He amends the rules by adding more principles which employ his notion of entrenchment. I have no suggestions for altering the forms of the principles which Goodman presents in Ch. IV of *Fact, Fiction, and Forecast*. I would alter their content by having them refer to the contrast between predicates which are not disjunctive and those which are, rather than to Goodman's contrast between well entrenched predicates and those less well entrenched. I believe that the principles so reformulated both rule out the same unacceptable

inferences for which the original principles were designed and do not rule out any inferences we want to retain.

Some think that we can imagine circumstances in which Goodman's principles would rule out inferences we should want to accept in favor of inferences we should want not to accept. I hope that my principles are less vulnerable to such criticisms. My main reason for advancing the principles, however, is not that I have objections to Goodman's line of attack but rather that the notion of disjunctiveness seems to be more basic than the notion of entrenchment. A predicate may be poorly entrenched because it is disjunctive, but, if I am right, a predicate is not disjunctive because it is poorly entrenched.

So far as I can see, the plausibility of the claims I have made about "grue" and "green" does not depend on the fact that "green" and "blue" are more familiar or, in Goodman's sense, better entrenched than "grue" and "bleen." The difference between disjunctive and nondisjunctive predicates is characterized by reference to the relations between the borderline cases of subpredicates with each other and with the borderline cases of the predicate of which they are subpredicates. Neither whether a predicate can be partitioned or split into subpredicates, nor whether something counts as a borderline case of a predicate, depends on the familiarity or entrenchment of the predicates in question. Relatively unfamiliar subpredicates, e.g., "contains vitamin B_1 " and "contains vitamin B_2 ," can be introduced to show that a better entrenched predicate, e.g., "contains vitamin B ," is disjunctive.

Perhaps we can imagine people who think that the correct use of "grue," but not the correct use of "green," is teachable by ostensive definition. Even if we can, this is irrelevant to my distinction, which has no direct connection with the possibility of ostensive teaching. For all I know, the uses of some disjunctive predicates can be taught ostensively without teaching the uses of the corresponding subpredicates. And the uses of some nondisjunctive predicates, e.g., "has a half-life less than one-thousandth of a second," cannot be taught ostensively.

Suppose some people have a term "vermson" which is coextensive with our "red or green." When they learn our use of "red," they regard the predicate "red" as disjunctive. For they call red things which tend toward orange "strong vermson," other red things "weak vermson," and green

⁷ *Fact, Fiction, and Forecast* (Cambridge, Mass., 1955), p. 67.

things "medium vermson." Perhaps this supposition is intelligible and one could elaborate the story to give a rationale for their use of "weak" and "strong." Even so, "vermson" is disjoint, and "red" is not. The subpredicates "weak or strong vermson" and "medium vermson" into which "vermson" can be partitioned, have no borderline cases in common. And the subpredicates "weak vermson" and "strong vermson," into which "red" can be partitioned, do have borderline cases in common. A further supposition that the people who naturally use "vermson" systematically disagree with us about borderline cases is inconsistent with our supposi-

tions about the coextensiveness of their predicates with ours. A disjoint predicate, e.g., one equivalent to "green and examined before t_0 , or red and not examined before t_0 " would have served Goodman's purposes as well as "grue." "Grue" and "bleen" happen to form an euphonious pair; and if Goodman had used a pair of disjoint predicates instead, we might have overlooked the subtler forms of exclusive disjunctiveness. Although "grue" is more difficult to treat than a disjoint predicate, I see no reason to suppose that its disjunctiveness is relative to the language we happen to speak or to the inferences of our forebears.

Dartmouth College

Received December 12, 1968

BOOKS RECEIVED

- BECK, Robert N., (ed.), *Perspectives in Philosophy*. New York: Holt, Rinehart and Winston, Inc., 1969. Pp. xv+524. Paper, \$5.50.
- BERKOVITS, Eliezer, *Man and God*. Detroit: Wayne State University Press, 1969. Pp. 376. \$12.50.
- BERNARDO, Giuliano di, *Logica, Norme, Azione*. Trento: Istituto Superiore di Scienze Sociali, 1969. Pp. 174.
- CALLAHAN, John F., *Augustine and the Greek Philosophers*. The Saint Augustine Lecture 1964. Villanova, Pa.: Villanova University Press, 1967. Pp. 117. \$3.50.
- CHATTERJI, P. C., *Fundamental Questions in Aesthetics*. Simla, India: Indian Institute of Advanced Study, 1968. Pp. xiii+202. \$8.00.
- COHEN, Robert S. and WARTOFSKY, Marx W., (eds.), *Boston Studies in the Philosophy of Science*, vol. IV, 1966-1968. Dordrecht, Holland: D. Reidel Publishing Company, 1969. Pp. 482.
- DAVIS, J. W., HOCKNEY, D. J., and WILSON, W. K., (eds.), *Philosophical Logic*. Dordrecht, Holland: D. Reidel Publishing Company, 1969. Pp. 277.
- DI GIANDOMENICO, MAURO, *Filosofia E. Medicina Sperimentale in Claude Bernard*. Bari: Adriatica Editrice, 1968. Pp. 303.
- EDIE, James E., (ed.), *New Essays in Phenomenology*. Chicago: Quadrangle Books, 1969. Pp. 383.
- FRANCESCHI, Alfredo, *Escritos Filosóficos*. Pp. 153.
- GANGULY, Sachindranath, *Wittgenstein's Tractatus: A preliminary*. West Bengal, India: Centre of Advanced Study in Philosophy, 1968. Pp. viii+129.
- GILL, Jerry H., (ed.), *Essays on Kierkegaard*. Minneapolis: Burgess Publishing Company, 1969. Pp. 197.
- HARRIS, Etrol E., *Fundamentals of Philosophy: A Study of a Classical Texts*. New York: Holt, Rinehart and Winston, Inc., 1969. Pp. 565. \$7.95.
- HARRISON, Frank R., III, *Deductive Logic and Descriptive Language*. Englewood Cliffs, New Jersey: Prentice-Hall, 1969. Pp. 534.
- HEILBRONER, Robert L., (ed.), *Economic Means and Social Ends: Essays in Political Economics*. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1969. Pp. 204.
- KAHANE, Howard, *Logic and Philosophy*. Belmont, California: Wadsworth Publishing Company, Inc., 1969. Pp. 441.
- KAMENKA, Eugene, *Marxism and Ethics*. New York: St. Martin's Press, 1969. Pp. 72. Paper, \$1.95.
- KRIMERMAN, Leonard I., (ed.), *The Nature & Scope of Social Science: A Critical Anthology*. New York: Appleton-Century-Crofts, Meredith Corporation, 1969. Pp. 796.
- KRISHNA, Daya, *Social Philosophy: Past and Future*. Simla, India: Indian Institute of Advanced Study, 1969. Pp. 82. \$2.70.
- KURTZ, Paul, (ed.), *Moral Problems in a Contemporary Society: Essays in Humanistic Ethics*. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1969. Pp. 301. \$6.95.
- LAPOINTE, Roger, *Consultation internationale sur le non-être*. Bruges, Paris: Desclée de Brouwer, 1969. Pp. 159.
- MICALLEF, John, *Philosophy of Existence*. New York: Philosophical Library, 1969. Pp. 235. \$6.50.
- MICHALOS, Alex C., *Principles of Logic*. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1969. Pp. 433. \$7.95.
- NEWMAN, Robert P. and NEWMAN, Dale R., *Evidence*. Boston: Houghton Mifflin Company, 1969. Pp. 246.
- O'CONNOR, John, (ed.), *Modern Materialism: Readings on Mind-Body Identity*. New York: Harcourt, Brace & World, Inc., 1969. Pp. 289. Paper, \$3.50.
- PATZIG, Günther, *Aristotle's Theory of the Syllogism*. Dordrecht, Holland: D. Reidel Publishing Company, 1969. Pp. 215. \$12.60.
- PIERCE, Donald John, *The Nature of History*. Ottawa, Canada: 1969. Pp. 75.
- SELLARS, Roy Wood, *Evolutionary Naturalism*. New York: Russell & Russell, 1969. Pp. 357.

- SMITH, James M. and SOSA, Ernest, (eds.), *Mill's Utilitarianism*. Belmont, California: Wadsworth Publishing Company, Inc., 1969. Pp. 177.
- STEGMÜLLER, Wolfgang, *Wissenschaftliche Erklärung und Begründung*. New York: Springer-Verlag, 1969. Pp. 811.
- TUCKER, John, *Lectures on the Foundations of Mathematics*. 1967.
- WALSH, W. H., *Hegelian Ethics*. New York: St. Martin's Press, 1969. Pp. 84. Paper, \$1.95.
- WARNOCK, G. J., *English Philosophy Since 1900*. Oxford: Oxford University Press, 1969. (paper). \$1.50
- WATANABE, Satoru, *Knowing and Guessing: A Quantitative Study of Inference and Information*. New York: John Wiley & Sons, Inc., 1969. Pp. 592.
- WESTPHAL, Fred A., *The Activity of Philosophy: A Concise Introduction*. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1969. Pp. 259.
- WISDOM, John, *Logical Constructions*. New York: Random House, 1969. Pp. 181. \$2.45.

**American Philosophical Quarterly
Monograph No. 4**

**Studies in the Theory of
Knowledge**

Edited by N. RESCHER

Contributions are by: Norman Malcolm, W. Donald Oliver, Peter Unger, John L. Pollock, John Knox, Jr., Alan R. White, Frederick Stoutland.

o 631 11480 4

35s. (£1.75) net

Jowett Papers, 1968-9

Edited by B. Y. KHANBAI, R. S. KATZ and A. PINEAU

The Jowett Society provides a focus for philosophical debate in the University of Oxford. This is a collection of some of the more notable contributions delivered in the past year, edited by three recent presidents of the society. Contributors are: A. N. Prior, A. W. Muller, J. R. Lucas, R. S. Katz, R. M. Hare, J. Kovesi, N. H. Dent, J. W. Yolton, A. Price, J. M. Finnis, Philippa Foot.

o 631 12450 0

About 25s. (£1.25) net

*** Philosophische
Grammatik**

LUDWIG WITTGENSTEIN

Edited by RUSH RHEES

Wittgenstein wrote this work during 1932-34—the period just before he began to dictate the Blue Book. In Part I he discusses the notions of 'proposition', 'sense', 'language', 'grammar'; in Part II he writes on logical inference, generality, leading to his discussion of mathematics, which fills four-fifths of the volume.

o 631 12350 4

63s. (£3.15) net

*** The Logic of Power**

INGMAR PÖRN

This essay has been written in the conviction that the methods of formal logic can be profitably used in the organization of social, political and legal thought. The author employs the basic ideas and techniques of current modal logic in the construction of a general theory of action and interaction.

o 631 12510 8

About 20s. (£1.00) net

*** Sacra Doctrina**

Reason and Revelation in Aquinas

PER ERIK PERSSON

Translated by J. A. ROSS MACKENZIE

The author focuses interest on Aquinas' *sacra doctrina* (his own term for what we would now call theology) and investigates the relationship of his ideas of revelation and of scripture with the elements of his thinking unmistakably derived from Greek philosophy.

o 631 11860 8

50s. (£2.50) net

*** Metaphysics and the
Philosophy of Science**

The Classical Origins: Descartes to Kant

GERD BUCHDAHL

Developments in science and its methodology are shown to have a profound influence on the classical philosophical systems, and the point of view applied brings to life the main philosophical figures with new significance.

o 631 11720 2

105s. (£5.25) net

*** The Methodological
Heritage of Newton**

**Proceedings of the University of
Western Ontario
Second Philosophy Colloquium**

Edited by R. E. BUTTS and J. W. DAVIS

The contributions are by an outstanding group of experts on this phase of the history of scientific thought: N. R. Hanson, F. E. L. Priestley, J. W. Davis, Gerd Buchdahl, L. L. Laudan, R. E. Butts, P. K. Feyerabend.

o 631 12200 1

35s. (£1.75) net

* Published separately in U.S.A.



**BASIL BLACKWELL
Oxford, England**

AMERICAN PHILOSOPHICAL QUARTERLY

MONOGRAPH SERIES

Edited by NICHOLAS RESCHER

The *American Philosophical Quarterly* also publishes a supplementary Monograph Series. Apart from the publication of original scholarly monographs, this is to include occasional collections of original articles on a common theme. The current monograph is included in the subscription, and past monographs can be purchased separately but are available to individual subscribers at *half price* (though not to institutional subscribers).

No. 1. STUDIES IN MORAL PHILOSOPHY. *Contents:* Kai Nielsen, "On Moral Truth"; Jesse Kalin, "On Ethical Egoism"; G. P. Henderson, "Moral Nihilism"; Michael Stocker, "Supererogation and Duties"; Lawrence Haworth, "Utility and Rights"; David Braybrooke, "Let Needs Diminish That Preferences May Prosper"; and Jerome B. Schneewind, "Whewell's Ethics." 1968, \$6.00.

No. 2. STUDIES IN LOGICAL THEORY. *Contents:* Montgomery Furth, "Two Types of Denotation"; Jaakko Hintikka, "Language-Games for Quantifiers"; James W. Cornman, "Types, Categories, and Nonsense"; Robert C. Stalnaker, "A Theory of Conditionals"; Alan Hausman and Charles Echelbarger, "Goodman's Nominalism"; Ted Honderich, "Truth: Austin, Strawson, Warnock"; and Colwyn Williamson, "Propositions and Abstract Propositions." 1968, \$6.00.

No. 3. STUDIES IN THE PHILOSOPHY OF SCIENCE. *Contents:* Peter Achinstein, "Explanation"; Keith Lehrer, "Theoretical Terms and Inductive Inference"; Lawrence Sklar, "The Conventionality of Geometry"; Mario Bunge, "What Are Physical Theories?"; B. R. Grunstra, "The Plausibility of the Entrenchment Concept"; Simon Blackburn, "Goodman's Paradox"; Stephen Spielman, "Assuming, Ascertaining, and Inductive Probability"; Joseph Agassi, "Popper on Learning from Experience"; D. H. Mellor, "Physics and Furniture"; and Michael Slote, "Religion, Science, and the Extraordinary." 1969, \$6.00.

AMERICAN PHILOSOPHICAL QUARTERLY

Edited by

NICHOLAS RESCHER

With the advice and assistance of the Board of Editorial Consultants:

Virgil C. Aldrich

Alan R. Anderson

Kurt Baier

Stephen F. Barker

Monroe Beardsley

Nuel D. Belnap, Jr.

Roderick M. Chisholm

L. Jonathan Cohen

James Collins

Arthur C. Danto

James M. Edie

José Ferrater-Mora

Richard M. Gale

Peter Thomas Geach

Adolf Grünbaum

Carl G. Hempel

John Hospers

Raymond Klibansky

Hugues Leblanc

Ernan McMullin

Benson Mates

John A. Passmore

Richard H. Popkin

Richard Rorty

George A. Schrader

Michael Scriven

Wilfrid Sellars

Alexander Sesonske

Manley H. Thompson, Jr.

John W. Yolton

VOLUME 7/NUMBER 3

JULY 1970

CONTENTS

- | | | | |
|--|-----|---|-----|
| I. JONATHAN BENNETT: <i>The Difference Between Right and Left</i> . . . | 175 | VI. NICHOLAS WOLTERSTORFF: <i>Objections to Predicative Relations</i> . . . | 238 |
| II. RICHARD RORTY: <i>Wittgenstein, Privileged Access, and Incommunicability</i> . . . | 192 | VII. JOSEPH BEATTY: <i>Forgiveness</i> . . . | 246 |
| III. RICHARD GALE: <i>Negative Statements</i> . . . | 206 | VIII. REX MARTIN: <i>On The Logic of Justifying Legal Punishment</i> . . . | 253 |
| IV. T. Y. HENDERSON: <i>In Defense of Thrasymachus</i> . . . | 218 | IX. JOSEPH MARGOLIS: <i>Egoism and the Confirmation of Metamoral Theories</i> . . . | 260 |
| V. HENRY E. KYBURG: <i>On A Certain Form of Philosophical Argument</i> . . . | 229 | <i>Books Received</i> . . . | 267 |

AMERICAN PHILOSOPHICAL QUARTERLY

POLICY

The *American Philosophical Quarterly* welcomes articles by philosophers of any country on any aspect of philosophy, substantive or historical. However, only self-sufficient articles will be published, and not news items, book reviews, critical notices, or "discussion notes."

MANUSCRIPTS

Contributions may be as short as 2,000 words or as long as 25,000. All manuscripts should be typewritten with wide margins, and at least double spacing between the lines. Footnotes should be used sparingly and should be numbered consecutively. They should also be typed with wide margins and double spacing. The original copy, not a carbon, should be submitted; authors should always retain at least one copy of their articles.

COMMUNICATIONS

Articles for publication, and all other editorial communications and enquiries, should be addressed to: The Editor, *American Philosophical Quarterly*, Department of Philosophy, University of Pittsburgh, Pittsburgh, Pennsylvania 15213.

OFFPRINTS

Authors will receive gratis two copies of the issue containing their contribution. Offprints can be purchased through arrangements made when checking proof. They will be charged for as follows: The first 50 offprints of 4 pages (or fraction thereof) cost \$12, increasing by \$1 for each additional 4 pages. Additional groups of 50 offprints of 4 pages cost \$8, increasing by \$1 for each additional 4 pages. Covers will be provided for offprints at a cost of \$4 per group of 50.

SUBSCRIPTIONS

The price *per annum* is eight dollars for individual subscribers and fourteen dollars for institutions. Checks and money orders should be made payable to the *American Philosophical Quarterly*. All back issues are available and are sold at the rate of three dollars to individuals, and four dollars to institutions. All correspondence regarding subscriptions and back orders should be addressed directly to the publisher (Basil Blackwell, Broad Street, Oxford, England).

MONOGRAPH SERIES

The *American Philosophical Quarterly* also publishes a supplementary Monograph Series. Apart from the publication of original scholarly monographs, this includes occasional collections of original articles on a common theme. The current monograph is included in the subscription, and past monographs can be purchased separately but are available to individual subscribers at a substantially reduced price. The back cover of the journal may be consulted for details.



I. THE DIFFERENCE BETWEEN RIGHT AND LEFT

JONATHAN BENNETT

I. THE "PARADOX" ABOUT RIGHT AND LEFT

KANT seems to have been the first to notice that there is something peculiar about the difference between right and left, but he failed to say exactly what the peculiarity is. His clearest account of the matter is in his inaugural lecture:¹

We cannot describe [in general terms] the distinction in a given space between things which lie towards one quarter, and things which are turned towards the opposite quarter. Thus if we take solids which are completely equal and similar but incongruent, such as the right and left hands . . . although in every respect which admits of being stated in terms intelligible to the mind through a verbal description they can be substituted for one another, there is yet a diversity which makes it impossible for their boundaries to coincide. (15 C.)

One can see roughly what Kant's point is. Take two coins which differ only in their spatial positions: any description of one in general terms also fits the other; but then it is also true that "their boundaries coincide" or, as Kant says elsewhere, that "each can be replaced by the other in all cases and all respects, without the exchange causing the slightest recognizable difference." For example, if I tell you that I earned *this* coin and stole *that*, then shuffle them and show them to you again, you cannot re-identify the one I earned unless you have tracked one of them through the shuffle.

A left and a right hand are more different than this. If I showed you two detached hands which differed only as right and left, told you that I was given *this* one and stole *that*, then shuffled and reproduced them, you could re-identify the stolen one without having tracked either through the shuffle. The two hands would be qualitatively different as well as numerically distinct; it would not be true that "each can be replaced by the other . . . without the exchange causing the slightest recognizable difference"; for example, a glove which fitted one would not fit the other. And yet, Kant thinks, this difference between the two hands cannot be "stated in terms intelligible to the mind through a

verbal description": he says that it is a qualitative difference which cannot be captured in language.

That is false. We can state in language what the difference is between the two hands, for we can describe one as "a right hand" and the other as "a left hand." If you did not see them before the shuffle, you can still identify the stolen one if someone tells you "The right hand is the one he stole"—this being meant and understood not as saying what arm the hand used to grow on but rather as describing the hand itself, as saying what *kind* of hand it is.

That refutes what Kant says, taken dead literally. Behind what he actually says, though, there is a less vulnerable claim about the meanings of "left" and "right" and their equivalents in other languages. It is the claim that one could explain the meanings of these words only by a kind of showing—one could not do it by telling. That is the claim I am going to explore.

Still, there was a point in skirmishing with Kant on the basis of a ploddingly literal reading of his words. He uses several unsatisfactory formulations like the one I have attacked, and these help him to think he can report something surprising—in one place he calls it a "paradox"—about the right/left distinction. "Two things can differ qualitatively, although the difference cannot be expressed in words"—that would indeed be surprising if it were not false! Again, Kant says that two hands whose boundaries do not coincide may nevertheless be "completely equal and similar," which would be astonishing if it were true. Other writers, too, have offered one-sentence formulations of what they suppose to be obviously a "problem" about right and left. Reichenbach, for instance, refers to "the problem of the existence of equal and similarly shaped figures that cannot be superimposed" (p. 109; see also Caird, p. 166). Taking "similarly shaped" to mean "having the same shape," that would be a problem indeed; but if you think that your hands have the same shape, just try putting a glove first on one and then, without turning it inside out, on the other.

¹ See the Bibliography at the end of the paper.

When Kant and others say that a left and a right hand have "similar shapes" or the like, perhaps they mean—as any mathematician would mean—that the hands do not differ in shape *except* to the extent that one is a right hand and the other a left. Then what they say is true. But now where is the "problem" or "paradox"? My two hands differ only as right and left; but they do differ in that way, so of course a single glove won't fit both. Why should I find this surprising or paradoxical?

There is indeed a peculiarity about the right/left distinction. But it does not lie on the surface: philosophical work will be needed to dig it out and lay it bare, and so it could not possibly afford a simple, immediate surprise of the sort Kant thought he had in store for us.

II. KANT'S USES OF THE "PARADOX"

The real peculiarity of the left/right distinction, as well as being more elusive than Kant realized, has a different kind of philosophical interest from any that he found in it. He tried to argue from it to some of his larger philosophical views, but without much success. His major attempt of this kind is definitively treated in a paper by Remnant; his minor ones are hardly worth discussing. Since Kemp Smith has fully described the roles which the left/right matter plays throughout Kant's writings (pp. 161–166), I need only to sketch them. This section and the next are not presupposed by the rest of the paper.

The relevant background facts are these. (1) Kant was a transcendental idealist, i.e., he held a certain view about the analysis of spatial concepts—*any* spatial concepts. (2) He took sides in the dispute about absolute versus relative space, i.e., the dispute about whether the concept of spatial location is more or less basic than that of spatial relations between things. (1) concerns the analysis of the basic spatial concepts, whatever they may be; whereas (2) concerns which spatial concepts are basic, whatever their further analysis might be. Yet we are told by Weyl:

Kant finds the clue to the riddle of left and right in transcendental idealism; (P. 84.)

and by Russell:

Right and left hands, spherical triangles, etc. . . . show, as Kant intended them to show, the essential relativity of space; (§150.)

and by Smart:

Kant supported the absolute theory of space. In particular he thought that the relational theory could not do justice to the difference between a left hand and a right hand. (P. 6.)

These conflicting accounts of Kant's intentions reflect the instability of those intentions themselves. Kant's first discussion of left and right was in a little paper in 1768. He returned to the topic briefly in his inaugural lecture of 1770. In the first *Critique* in 1781 he took over much of that lecture almost verbatim, but made no mention of left and right. In the *Prolegomena* of 1783, intended as a popular summary of the *Critique*, he resurrected left/right and gave it a short section to itself. But then in the second edition of the *Critique* (1787), in which several new arguments and emphases are borrowed from the *Prolegomena*, the left/right matter once more disappears from sight. Kant seems to have been genuinely unsure whether he could draw philosophical conclusions from his point about the right/left distinction.

He also wavered in his views about *what* conclusions he could draw. Although he did not firmly enough distinguish (1) the issue over transcendental idealism from (2) the issue about absolute versus relative space, it is not too misleading to say: in 1768 he used the left/right matter to support the absolute theory of space; in 1783 he took it to support transcendental idealism; while in 1770 he adduced it in support of a doctrine which is not quite either of these though it arguably entails both.

In short, Kant could not decide which if any of his doctrines about space can draw strength from special facts about the right/left distinction. I am sure none of them can.

III. PROLEGOMENA § 13

Behind Kant's words in the inaugural lecture I have detected the claim that an explanation of the meanings of "right" and "left" requires *showing*, i.e., demands an appeal to sensorily presented examples. I shall call this claim the Kantian Hypothesis. It may not be what Kant "really meant" when he wrote about right and left, but it is the best we can get from him. In defense of this contention I shall examine *Prolegomena* §13, which is Kant's longest and most detailed treatment of the matter, and also, I believe, his last. When examined carefully, this passage can be seen to amount to a series of pointers toward the Kantian Hypothesis. This is not a bad thing to amount to; and really my only criticism is that in *Prolegomena* §13 Kant

purports to be expressing, not merely pointing toward, the peculiarity of the right/left distinction. (In the final sentence I make two corrections which the translator accepts. The numbers are for subsequent reference.)

[One would have thought that] if two things are [1] completely the same in all points that can be known at all about each separately (in all determinations belonging to quantity and quality), it must follow that each can be replaced by the other in all cases and all respects, without the exchange causing the slightest recognizable difference. This is in fact the case with plane figures in geometry; but various spherical figures show, notwithstanding this [2] complete inner agreement, an outer relation such that one cannot be replaced by the other. For example two spherical triangles on opposite hemispheres which have an arc of the equator as their common base can be completely equal, in respect of sides as well as angles, so that [3] nothing is found in either, when it is described alone and completely, which does not also appear in the description of the other (on the opposite hemisphere). Here then is an *inner* difference between the two triangles which [4] no understanding can show to be inner and which only reveals itself through the outer relation in space. But I will quote more usual cases which can be taken from ordinary life.

What can be more like my hand or my ear, and more equal in all points, than its image in the mirror? And yet I cannot put such a hand as is seen in the mirror in the place of its original: for if the original was a right hand, the hand in the mirror is a left hand, and the image of a right ear is a left ear, which could never serve as a substitute for the other. Here are [5] no inner differences that any understanding could think; and yet the differences are inner so far as the senses tell us, for the left hand cannot be enclosed in the same boundaries as the right (they cannot be congruent) notwithstanding all their mutual equality and similarity; the glove of one hand cannot be used on the other. . . . We cannot make the difference between similar and equal but yet incongruent things (e.g. spirals winding opposite ways) [6] intelligible by any concept whatsoever, but only by their relation to the right and left hand, which immediately involves intuition.

I have omitted Kant's "solution." The question I want to answer is: What is his problem?

The problem, as Kant sees it, is that a certain plausible proposition is false. (My addition of "One would have thought that" at the start of the passage, though it wrongly makes Kant explicit about this, must be legitimate. Without it, Kant asserts something which he immediately proceeds to deny.) The proposition in question has the form

$$(x)(y)(Fxy \rightarrow Gxy).$$

Kant says that, although this is plausible, there are in fact values of x and y such that $(Fxy \ \& \ \sim Gxy)$; and to solve his problem will be to explain this surprising fact. Our problem is to discover what F and G are.

There is no difficulty about G . Gxy is the statement that x can be replaced by y "without the exchange causing the slightest recognizable difference." Thus Gxy is true if x and y are newly minted coins from the same die, and false if they are a normal pair of hands, i.e., a pair differing only as right and left.

The search for F is embodied in the question: What does Kant think he can say about a normal pair of hands from which one might naturally, though wrongly, infer that they could not be told apart? We can safely pin everything on the one example of a pair of hands, for it is universally agreed that in this area Kant's examples stand or fall together.

He expresses Fxy in six different ways. Here are two of them:

- (1) x and y "are completely the same in all points that can be known at all about each separately (in all determinations belonging to quantity and quality)."
- (3) When x is "described alone and completely," its description is the same as y 's.

To describe something "alone and completely" is presumably to say everything about it except how it relates—spatially and otherwise—to other things. But then is (3) true of a normal pair of hands? In describing one of the pair "completely" we can use a phrase which does not fit the other, namely "a right hand"—taking this to express a fact not about which arm it grows on but about its shape, e.g., about which sort of glove will fit it. To exclude this, Kant must say that if we use "right" in describing a hand we are not describing it "alone": the phrase "a right hand," he must say, is covertly relational, and not merely in the attenuated way in which any description, e.g., "a small hand," is covertly relational. This is not obviously true, and the only arguments I can find to support it stem from the Kantian Hypothesis.

If (1) is not also to amount to a pointer toward the Kantian Hypothesis, the phrase "determinations [= properties] belonging to quantity and quality" must be turned to account. But it cannot be. The difference between a left and a right hand is "qualitative" in any plain sense of the word; and Kant's technical sense of "quality" in the *Critique* is too unclear to help us here.

Here are Kant's other four ways of expressing *Fxy*:

- (2) There is a "complete inner agreement" between *x* and *y*.
- (4) The "inner difference" between *x* and *y* is one which "no understanding can show to be inner."
- (5) Between *x* and *y* there are "no inner differences that any understanding could think."
- (6) "We cannot make the difference between [*x* and *y*] intelligible by any concept whatsoever."

We must presume (2) to be a careless contraction of (4) or (5). Otherwise, Kant is saying that between *x* and *y* there is (2) a "complete inner agreement" and also (4) an "inner difference." So (2) can be ignored.

(5) and (6) go together. For Kant, "the understanding" is the faculty of "concepts": to be thought by the understanding is to be brought under, thought through, or made intelligible by, concepts. So (5) and (6) both say that a right hand need not fall under any concepts which do not equally apply to a left hand, which is tantamount to denying that there is any concept of rightness-as-distinct-from-leftness. Since "right as distinct from left" is a meaningful description, why should Kant deny that there is a concept corresponding to it? His only hint at an answer is in his remark, at the end of the passage, that we can explain the right/left difference only in a way "which immediately involves intuition [= sense-experience]." But this is—and so (5) and (6) are just unargued pointers toward—the Kantian Hypothesis.

Whereas (5) and (6) say that the understanding cannot show or express *what the left/right difference is*, (4) says that it cannot show *that the difference is an inner one*, implying that one could show this only with the aid of "intuition" or sense-experience. To assess this, we must know what an "inner" difference is. It seems to be just a difference in respect to something other than spatial location or orientation—a difference in respect to some property that a thing can carry around with it. This yields the wanted result that there is an inner difference between a pair of normal hands, and not between two new coins from the same die. It also fits my example in Sect. I above: if two things are to be separately re-identifiable after a shuffle, without being tracked through it, what is needed is precisely some "inner" difference between them,

i.e., some difference of the kind that can be carried through a shuffle.

So (4) seems to say that someone who has grasped what the difference is between a right and a left hand must make a further appeal to experience if he is to grasp that one hand cannot be made congruent with the other just by moving it around. This is in fact correct; for there are mathematically possible spaces in which a right hand could, by sheer travel, become a left hand; and if our space is not of such a kind, that is an empirical fact about it and in that sense a fact which can be known only by appeal to experience. But it is not credible that that is the point Kant was trying to make in (4). I am sure that what he says about showing (4) *that the difference is inner* is meant to follow from what he says about showing (5, 6) *what the difference is*. When he says at (5) in the quoted passage:

Here are no inner differences that any understanding could think; and yet the differences are inner so far as the senses tell us,

isn't it clear that he is simply failing to distinguish "what the inner difference is" from "that the difference is inner"? If he is, and if that explains (4), then the latter goes the same way as (5) and (6)—toward the Kantian Hypothesis.

So *Prolegomena* §13 does, to its great credit, yield the Kantian Hypothesis. But that is all it yields; and it does not make clear just what the force of the Hypothesis is, or why it is true. There remains work to be done.

Before we go on with it, there are two footnotes to the claim that Kant was the first philosopher to notice that right/left is peculiar.

In a letter to Clarke, Leibniz says that God could have no reason for choosing (a) the way things are in fact arranged in space rather than (b) an arrangement "preserving the same situation of bodies among themselves" and differing from (a) only in "changing East into West"; whence he infers that (a) and (b) are not really different (p. 26). He probably thinks of (b) as the world's being rotated through 180°, changing north into south as well as east into west. Still, all he actually says is "changing East into West"; so he could be envisaging a systematic left/right switch, or mirror-image transformation, in which case he has anticipated something like Kant's point. I find the latter reading implausible. It credits Leibniz with introducing an original philosophical insight in an incredibly offhand way, and arguing from it—without first explaining or defending it—even

though he could further his main argument much less vulnerably with the rigid-rotation version of (b). Also, when he reverts to this matter in his next letter to Clarke he clearly construes it in the rigid-rotation rather than the mirror-transformation manner (p. 37). Kant's thoughts about right and left, however, grew out of his disagreements with Leibniz, and the east/west remark may well be what put him on the track.

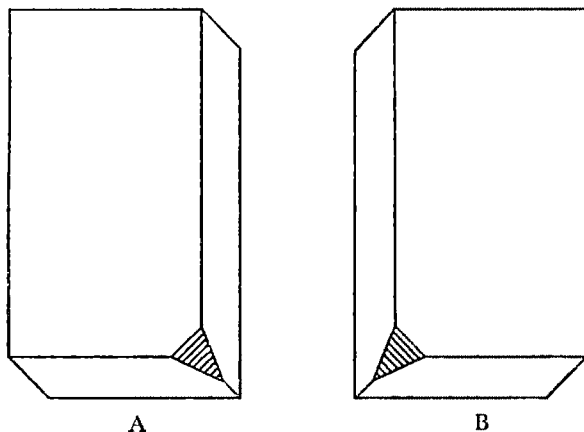
The 11th century Arab philosopher Ghazali has a better claim to have anticipated Kant's insight:

The highest sphere moves from east to west and the spheres beneath it in the opposite direction, but everything that happens in this way would happen equally if the reverse took place, i.e., if the highest sphere moved from west to east and the lower spheres in the opposite direction. For all the same differences in configuration would arise just as well. Granted that these movements are circular and in opposite directions, both directions are equivalent; why then is the one distinguished from the other, which is similar to it? (Quoted in Averroës, Vol. I, p. 30.)

(I am indebted to George F. Hourani for calling this passage to my attention.)

IV. ENANTIOMORPHISM

It is a nuisance that, when we want to use "a left hand" to mean something about the hand's shape, what sort of glove will fit it, etc., the phrase can also mean "a hand that grows on a left arm." In either meaning it applies to just the same objects, but that is a mere contingency. For this reason, and for others that will emerge shortly, hands are not the best example of the relationship we are interested in. I prefer these two boxes:



In Kantian language, these differ as "things which lie towards one quarter and things which are turned

toward the opposite quarter." Such pairs are sometimes called "incongruous counterparts," which means that (a) their boundaries do not coincide, and that (b) one of them looks just as the other would in a mirror. If the sliced-off corners were restored, (a) would be false and the boxes would not be "incongruous"; if just one had its corner restored, or if one were bigger, (b) would be false and the boxes would not be "counterparts."

The mathematical term for two things which are thus related is *enantiomorphs* ("having contrary shapes"). I shall sometimes use this word instead of the longer "incongruous counterparts," but not to mark any distinction.

It is time to confess that my paper's real topic is not right/left as such, but rather enantiomorphism, or the difference between incongruous counterparts. The right/left distinction can bear the whole weight of the difference between any pair of enantiomorphs: that is, any such pair can be so described that a "right"/"left" switch turns a description of either into a description of the other. In this section I shall show how such descriptions work, to show that in discussing incongruous counterparts it is *convenient but not essential* to use "right" and "left" or some other pair of terms which similarly refer to the two sides of the human body.

If the two boxes *A* and *B* are to be described by the use of "right" and "left," without anything's being assumed, it apparently cannot be done more simply than this:

A: When (1) the line from its small cut to its small uncut face runs the same way as the line from your feet to your head, and (2) the line from its large cut to its large uncut face runs the same way as the line from your back to your front, then (3) the line from its middling cut to its middling uncut face runs the same way as the line from your right side to your left side.

B: Switch "left" and "right" in the above description of *A*.

The following would be simpler, but they make assumptions:

A: When (1) its small cut face is downmost and (2) its large cut face is toward you, then (3) its middling cut face is to your right.

B: Replace "right" by "left" in the above description of *A*.

Those simpler versions are accurate if you are on your feet and facing the box, or on your head with your back to it. They are wrong if you are on your feet with your back to the box, or on your head facing it.

What the longer descriptions make explicit is that we use "right" and "left" to express the difference between an object and its incongruous counterpart by fixing directions along two of the object's dimensions and then employing "right" and "left" to make the required distinction in the third dimension. (Here and throughout I ignore the mathematically sound but entirely unhelpful remark—e.g., in Wittgenstein, 6.36111—that in a fourth spatial dimension *A* could be flipped over so as to become congruous with *B*.) To discriminate *A* from *B* by reference to the human body in this way, we need to be able to pick out three axes of the human body and to be able to distinguish the two directions along each axis. It is harder to distinguish directions along the left/right axis than along either the head/feet or the back/front axis; but this fact, which connects with our being broadly and superficially left/right symmetrical, is irrelevant to the use of human bodies to discriminate *A* from *B*. My first description of *B* above could just as accurately have ordered a "head"/"feet" or a "front"/"back" switch in the long description of *A*.

We can also use "right (side)" and "left (side)" to distinguish the two sorts of hand, and not through the contingency about which sort of hand grows on which side (I now use a self-explanatory shorthand).

Left hand: When thumb → little-finger runs with back → front, and wrist → fingertips runs with feet → head, then palm → knuckles runs with right-side → left-side.

Right hand: Switch "right" and "left" in the above description of the left hand.

But the two sorts of hand can be distinguished without reference to human flanks, just so long as we have some pair of enantiomorphs—e.g., the two boxes—to use as a standard:

Left hand: When thumb → little-finger runs with large-cut → large-uncut face of *A*, and wrist → fingertips runs with small-cut → small-uncut face of *A*, then palm → knuckles runs with

middling-cut → middling-uncut face of *A*.

Right hand: Replace "*A*" by "*B*" in the above description of the left hand.

It is commonly believed that the distinction between a pair of enantiomorphs, when properly spelled out, must refer to the "point of view" of an "observer"; but this is false if it goes beyond the general point that any empirical distinction must, qua empirical, have a possible observer lurking in the conceptual background. The idea seems to be that we should describe *A* like this: "When the line from its small cut to its small uncut face runs the same way as the line from the observer's feet to his head . . . etc.". But if a human body is used in describing *A*, why should it be an observer's body? A corpse would serve as well.

In any case, human bodies are not needed at all. It is sometimes said that we can distinguish enantiomorphs only because our bodies are asymmetrical in at least two dimensions, but this is false too. If our bodies were symmetrical about a point we could still make the distinction we now make in terms of "right" and "left," the one exemplified by *A* and *B*; only we should have to express it in terms of something other than the sides of our bodies. Perhaps it is worth a paragraph to explain how this might be done.

Traveling from Ridge toward Loughed, I must turn left at a certain corner to reach the University. If humans were spherical I might be told which way to roll at that corner by reference to the box *A*:

If (1) small-cut → small-uncut face of *A* runs with ground → sky, and (2) large-cut → large-uncut face of *A* runs with turning-corner → Loughed, then (3) middling-cut → middling-uncut face of *A* runs with the next part of your journey.

That may seem to compare ill with the instructions I can in fact be given:

If at that corner you (1) stand (2) facing Loughed, (3) you must turn left before proceeding;

but this, though briefer, is not logically simpler. It spells out into:

If you so orientate yourself that (1) feet → head runs with ground → sky, and (2) back → front runs with turning-corner → Loughed, then (3) right-side → left-side runs with the next part of your journey.

Also, it is routine work to construct definitions of "A-turn" and "B-turn" which would let us describe a route unambiguously and quite briefly by specifying where the spherical traveler should make an A-turn and where a B-turn. I have heard it insisted that if our bodies were spherical we could not remember the difference between *A* and *B*, or between *A*-like boxes and *B*-like boxes, or between *A*-turns and *B*-turns; but I know of no principles in the epistemology of spherical rational animals which could justify this claim.

Failure to grasp the conventions underlying our use of "left" and "right" has generated the mildly famous "mirror problem": why does a mirror reverse left/right but not up/down? Martin Gardner (pp. 29-31) presents the only clear account I know of the solution to this: the answer to "Why does a mirror . . . etc.?" is *It doesn't!* Your image in a normal mirror is a visual representation of an incongruous counterpart of your body, and we conventionally describe this sort of relationship as a "left/right reversal." But this convention does not pick out one dimension as privileged over the other two: it is merely a natural and convenient way of expressing the fact of enantiomorphism in a case where each member of the enantiomorphic pair has—like a normal human body—a superficial over-all bilateral symmetry. (Of course an object which was precisely and totally bilaterally symmetrical could not have an enantiomorph.) If we are to describe what an ordinary mirror does, in a way which really does select one axis of the body in preference to the other two, then we must say this: if you face the mirror, it reverses you back/front; if you stand side-on to it, it reverses you left/right; if you stand on it, it reverses you up/down. These facts, once they are properly described, do not offer a problem. They are explained by routine optics. For some deeper aspects of this matter, see the paper by Pears.

V. "WHAT IS THE DIFFERENCE?"

I am going to test the Kantian Hypothesis that the difference between right and left—by which I really mean "the difference between anything and its enantiomorph"—can be explained only by showing and not by telling. Now, there is one way of taking this in which it is obviously false, the following being a counter-example:

If you have a man on one side of you and a woman on the other, then you have *either* a man

on your left and a woman on your right *or* a man on your right and a woman on your left, depending upon which side each is on.

Or we can tell someone what the difference is between the boxes *A* and *B* by giving him a mathematical description of each (the two descriptions will differ only in that one will have a minus-sign before each value for x), and telling him that of these two descriptions one fits *A* and the other fits *B*.

In ways like these we can explain *the difference*: we can say what distinction is marked by "right" and "left," or what kind of difference there is between a pair of incongruous counterparts, without saying anything about how to tell which is which. Analogously, someone might learn what "the difference between" blue and green is by being told that sunny skies characteristically have one of these colors and well-watered grass the other. Confronted with two shirts, say, he would then be in a position to say "I know what the difference between these is—one is blue and the other green"; but he would not be able to say which is blue and which green.

When Kant says—in episode (4) of the long passage—that between two incongruous counterparts there is "an inner difference which no understanding can show to be inner," he may mean that one could not explain in general terms "what the difference is" even in this attenuated sense. If so, he is surely wrong. (Thus Weyl, p. 80. But Weyl errs in thinking that this is Kant's only point.)

The Kantian Hypothesis that I want to discuss says that we must use sensorily presented instances—must resort to showing—if we are to explain the *direction* of the left/right distinction, i.e., to explain which is which. I shall for brevity's sake go on using the phrase "the difference between," but always intending it in this which-is-which manner. In my use, someone does not know the difference between right and left unless he knows which is his right side and which his left; and we have not told someone what the difference is between *A* and *B* unless we have equipped him to pick out *A* as distinct from *B*.

VI. TACTICS

A good way of examining how something could be explained is to consider how someone could discover that he has it wrong. So I shall invent someone—call him an Alphan—whose grasp of English is perfect except that he gives to "right" the

meaning of "left" and vice versa. We have to see how he could learn of his mistake.

For a contrast case I shall take someone—call him a Betan—whose grasp of English is perfect except that he has switched the meanings of some other pair of spatial expressions. The Betan's mistake concerns the word "between": he gives to the form " x is between y and z " the meaning we give to " y is between x and z ". (He thinks that the thing asserted to be between the other two is the thing whose name occurs between the names of the other two: any English sentence containing the form " x is between y and z " is a kind of *picture* of what the Betan thinks it means.) The contrasts I shall draw between the Alphan and the Betan will not depend at all upon special features of betweenness—e.g., that it is a triadic relation, or that it concerns order rather than shape or size. Essentially the same contrasts could be drawn if the Betan had switched the meanings of "large" and "small," "inside" and "outside," "round" and "square," or any one of dozens of other pairs of spatial expressions. Nor does it matter that the Betan has not switched a pair of *words*. Pretend that English also contains "between," defined by " y is between x and z " = " x is between y and z ," and then think of the Betan as having switched the meanings of "between" and "between."

Let us ask how the Alphan and the Betan can discover their respective semantic errors. In seeing how the two cases differ we shall see that the Kantian Hypothesis is nearly true.

If the Alphan encounters a statement using "right" or "left" which he knows to be false given the meanings he attaches to those words, but which might for all he knows be true if their meanings were switched, he may guess that the speaker or writer is mistaken or lying. As such cases pile up, however, the Alphan ought to conclude that *he* has made an error—a semantic one. Similarly, the Betan will realize his mistake about "between" if he encounters enough statements which he knows to be false on his understanding of them but which might for all he knows be true on the other relevant interpretation, i.e., the one which is in fact correct.

I shall take these to be the only ways in which either man can discover his error. Any corrective force that verbal definitions have can be expressed in the pattern of correction I have described, and it will make for clarity if everything is brought under the one pattern.

So our question about each man is: What true statements will he, interpreting them in his mistaken

way, think to be false? The inquiry is not a psychological one. The intellectual responses of the Alphan and the Betan are dramatic embodiments of logical relations, so we credit them both with maximum alertness, intelligence, retentiveness, and so on.

VII. ADMISSIBLE EVIDENCE

Here are some boring ways of correcting the Alphan. Say to him "I am now touching your right shoulder," while touching his right shoulder. Say to him "Your right shoulder is the one with the birthmark," when he knows which of his shoulders has a birthmark and it is indeed his right shoulder. Say to him "As I stood facing Boulogne, I had Dover on my left and Folkestone on my right," and give him a map of Europe or a look at Europe.

All these correct him by applying "right" and "left" to particular bits of the world of which he has relevant independent knowledge—from his own observation of those particulars, or from inspecting maps or pictures or statues of them. It is obvious—and the Kantian Hypothesis does not deny—that the Alphan can be corrected in ways like these, as indeed can the Betan. What the Hypothesis says, in effect, is that if we rigorously exclude all such references to particulars which are also known through observation, the Betan can still be corrected while the Alphan cannot. If we are to test the Hypothesis, therefore, we must deprive both men of any statements referring to particulars which they know about from observation.

We must also ban all English statements about particulars which the Alphan or Betan knows about from hearsay in languages other than English. Any attempt to capitalize on the Alphan's correct grasp of some pair of non-English synonyms of "right" and "left" would merely force us to redirect our inquiry—making us ask about his grasp of those other words rather than of "left" and "right"—without altering the inquiry's fundamental nature.

So the English statements encountered by the Alphan or Betan are to say *nothing relevant* about any particular things or places or situations regarding which he has *any relevant* knowledge from any source other than what he reads in English. The word "relevant" here means "relevant to his semantic mistake," and it isn't always clear whether something is relevant in this way. Rather than constantly watching for hidden relevances, let us exclude more: the English statements encountered by the Alphan or the Betan are to say *nothing at all*

about any particulars regarding which he knows *anything at all* from any source other than what he reads in English. This will be much easier to handle, and it cannot affect the validity of our results: anything allowed in by the weaker exclusion but kept out by the stronger must, *ex hypothesi*, be irrelevant to the matter in hand.

Think of each man as receiving an account, written in English, of some part of reality about which he knows nothing from any other source (and, in the meantime, forget that this involves his receiving ink-samples about which his correspondent might make comments in English). It is crucial that they are to know *nothing* about the described part of reality other than (a) general truths about it which hold true of all reality, and (b) truths about it in particular, or about particulars in it, which they learn simply from what they read in English. They can be in a position to say of something they observe, "This is *a thing of the kind* the Englishman was referring to when he wrote . . .," but never to say "This is *the thing* the Englishman was referring to when he wrote . . ." They must not even be in a position to relate particulars described to them in English with particulars presented in any other way, apart from merely comparing them. So they must not be in a position to say "This rain was caused by the atomic explosion the Englishman wrote about" or "The mountain the Englishman wrote about is 7,568 miles NNE of my village." It follows that among the things they must not know about the part of reality described to them in English is where it is in relation to themselves.

The line of exclusion I am drawing is not arbitrary or willful. There is a good reason for depriving both Alphan and Betan of any independent knowledge, however remote and relational, of any particular they read about in English. Everything thus excluded is either irrelevant to our inquiry or else logically on a continuum with the trivial case where we touch the Alphan's shoulder while saying "I am now touching your right shoulder."

Even with all this excluded, the Alphan and Betan can still encounter millions of uses "left" and "right," or of "between." And they may still be able to judge some of what they read to be false; for one can reject a statement about a particular of which one has no independent knowledge, on the grounds that it conflicts with a generalization which one knows to be true. I heard the BBC say that 9,000 civilians would be evacuated from Aden

within a year, at the rate of 500 per month: without investigating Aden I was entitled to reject *that*—the thing is logically impossible. In Shelley's *The Cenci*, a torturer says of his intended victim:

As soon as we
Had bound him on the wheel, he smiled on us,
As one who baffles a deep adversary;
And holding his breath, died.

I wasn't there; but I know that this report is false—Marzio cannot have committed suicide by holding his breath, because that is a physiological impossibility.

Of those two examples, one concerns a logical generalization, and the other a contingent, broadly causal generalization. I shall use this dichotomy in what follows.

VIII. LOGICAL CLUES: THE BETAN

There are countless "logical clues" to the Betan's error—that is, countless true statements which, interpreted according to his semantic error, will come out logically false. Here are two examples, with the Betan's pictures indicated in brackets:

- (a) "I sat between a silent old bore and a talkative young bore [I-bore-bore]. Since there were only two bores present, I resented having one on each side of me."
- (b) "Since Baltimore is between Washington and New York [B-W-NY], and we were flying in a straight line, we passed over New York first, then Baltimore, then Washington."

These bring the Betan's correct understanding of "each" and "side," and of "straight" and "first" and "then," into logical conflict with his incorrect reading of the form "*x* is between *y* and *z*." With no independent knowledge of the dinner or of the flight, he nevertheless knows that there is something amiss with each statement or with his understanding of it.

Those statements are logical clues for the Betan only because he does understand all the other words correctly. Perhaps, then, we can shield him from logical clues to his error about "between" by supposing that he errs also about other words such as "straight" and "each," and that these other errors *match* his mistake about "between." Can we do this? Can we credit him with a set of semantic errors which dovetail together so that no true statement will give him a logical clue to his having mis-

understood "between" or any of the other words in the set?

("Between" can conflict with itself, because the Betan would equate " x is between y and z " with " z is between y and x " but not with " x is between z and y ." But since that is a special feature of "between," and would not obtain for most of the examples I might have taken as contrasts to right/left, I cannot avail myself of it. The Betan is in enough trouble anyway.)

The first point to notice is that dozens of words have direct meaning-connections with "between." To remain sheltered from logical clues to his error about "between," the Betan must err about the meanings not just of the words I have mentioned but also of "symmetrical," "lopsided," "middle," "pinch," "trapped," "separated," and many more.

Also, it is hard to see *what* semantic error we must suppose him to make in each case. In (b), for instance, will he give to the sentence "We passed over New York first, then Baltimore, then Washington" the meaning we give to "We passed over Baltimore first, then Washington, then New York"? It is not clear what underlying semantic error, concerning what word(s) or phrase(s), could generate that reading of the sentence.

Finally, if he is to have no logical clue to any of his semantic errors, then each error with which we initially credit him will presumably have to be matched by yet others, these in their turn by others again, and so on outward. I can't illustrate this in detail because, as just noted, I can't say what semantic error is required in any single case; but I am sure that if we could specify a semantic error which would produce a "match" in a given case, it would be one which could remain unclued only if matched by further errors. For example, if we try to draw (a)'s sting by supposing the Betan to make a matching mistake about the word "each," then we must protect the latter mistake from statements which connect "each" with such words as "both" and "two" and "neither" and so on. The Betan's semantic errors, in short, must ramify until they infect his understanding of most words in the language—and far beyond the point where we could still say that he does, with certain exceptions, understand English.

The proposed revision in our account of the Betan is, therefore, impossible.

IX. LOGICAL CLUES: THE ALPHAN

What logical clues could the Alphan have to his

error about the meanings of "right" and "left"? That is, what true statements might he read which, on his interpretation of them, would be logically false? Perhaps these would do:

- (a) "As I stood on the deck facing forward, a gun to my right fired a short burst. It was the starboard oerlikon."
- (b) "As a pitcher he is a southpaw—he can't pitch at all with his right hand."

Confronted with either of these, the Alphan would smell a rat—provided he understood "starboard" and "southpaw" correctly.

Can we protect him from any such logical clues by crediting him with matching semantic errors?

It is encouraging that so few words are involved. Indeed, the only certain examples I can find—apart from ones drawn from very limited dialects—are "port" and "starboard," "southpaw," the words for the four points of the compass, and a few cricketing terms. Also "clockwise" and "anticlockwise," *if* it is contingent that most clock-hands move clockwise. I have doubtless missed some, but not many.

Still, the language could have been otherwise. Screws might be called "standard" and "non-standard" according to how they have to be rotated to be driven in, a righthanded golf club might be called "a hogan" and a lefthanded one "a charles," and so on. Let us pretend, as we easily can, that hundreds of English words are thus meaning-connected with the left/right distinction: *now* can we shield the Alphan from logical clues by the "matching errors" move?

Easily! In each case we know exactly what the matching semantic error must be, namely a simple switch—of the meanings of "port" and "starboard," "hogan" and "charles," and so on. Furthermore, these errors need not ramify and infect words which are not directly meaning-connected with "right" and "left." The initial set of switches completes the whole job, leaving the Alphan with no source of logical clues to his error about "right" and "left" or to any of his compensating semantic errors.

So we can, for example, comfortably suppose that he begins with his mistake about "right" and "left" and is smoothly seduced by it into his other mistakes without ever having, so far as meaning-relationships are concerned, the faintest hint that he has gone astray. The analogous supposition about the Betan collapses in chaos.

That, then, is my first contrast between "left"/"right" and "between"—indeed, I believe, between

"left"/"right" and any pair of spatial terms which is not equivalent to the left/right distinction. Our terminology for the left/right distinction, unlike any other part of our spatial terminology, has an extremely simple internal logical structure and is thoroughly insulated from the rest of the language. It is for those two reasons that good dictionaries, which do not define "between" as "the normal relation of the mouth to the nose and chin," or "round" as "the normal shape of the pupil of a human eye," do perforce define "right" in terms of "that hand which is normally the stronger of the two."

X. CONTINGENT CLUES: THE BETAN

I now drop logical clues to ask what "contingent clues" either of our men could have to his semantic error. That is, what true statements can he read which, interpreted as he will interpret them, conflict with contingent generalizations which he knows to be true?

Here are some contingent clues for the Betan, again with his pictures indicated in brackets:

- (a) "James stood between a snow-clad mountain and me [James-mountain-me]: I could see him perfectly."
- (b) "Finding myself between a sheer cliff and the oncoming tide [me-cliff-tide], I was naturally afraid that I should be drowned."
- (c) "My brother flung himself between the gun and my body [brother-gun-me], so that the bullet hit him instead of me."

Let us see whether the Betan can evade the force of all such contingent clues, in the following way. Each time he reads a statement which, on his understanding of it, conflicts with a generalization which he has hitherto accepted, he concludes that the generalization does not hold true in the part of the world described in the statement (call it "England"). This would enable him to think that the statement is true on his interpretation of it, and is therefore not evidence that he has made a semantic mistake. It does not matter that he would be silly to try to neutralize each contingent clue by supposing that in England things happen differently. My question is: Can he succeed?

Well, under this strategy he must suppose that in England (a) things can be seen through snow-covered mountains, (b) the sea can scale sheer cliffs, and (c) bullets can swerve without being physically deflected. Furthermore, as clues accumulate, and

as some occur containing "because," "since," "so," etc., the Betan must suppose that these strange things which can happen in England do regularly happen there in certain conditions: in England (a) an intervening snow-clad mountain *improves* one's view of dark objects beyond it, (b) waves *are drawn* up sheer cliffs by people at the top, (c) the availability of an alternative target *turns* a bullet in its tracks.

After dealing with variants of just three statements, the Betan already has a strange picture of English life; but there is worse to come. For one thing, each of his suppositions must be reconciled with the rest of what he reads about England, and this will force him into other, equally wild suppositions. (False factual beliefs, indeed, may not suffice. For our rules allow him to read such statements as "In England waves are *not* drawn up cliffs by the presence of people at the top," which would require him to make a semantic error about—of all words—"not.") And those three examples plus their progeny are only a tiny fragment of all the contingent clues he can encounter. There will be others, involving thousands of familiar, fundamental aspects of the behavior of the macroscopic world; and each will require him to think that England is different in the relevant respect and in hosts of other respects which follow from that.

If the Betan executes even a small portion of this clue-canceling strategy, he will lose control of his picture of how things happen in England: it is humanly impossible to go any distance with this strategy. To take it all the way, however, is not just psychologically but logically impossible for the Betan as we have described him; for if we suppose him to adopt, remember, and retain all the beliefs about England demanded by his strategy, we must retract our original stipulation that he does, in the main, understand the English language. For example, we cannot say that he knows what "bullet" means if he has endless false beliefs about how the things properly called "bullets" behave, what they look like, what their structure is, and so on. Yet the proposed strategy, if applied to a suitable range of contingent clues, will indeed leave the Betan with hardly any true beliefs about bullets: when shown a real bullet he certainly won't classify it as an object of the sort called "bullet" in English, and the longer he studies it the less inclined he will be to classify it thus. But this is just to say that he doesn't know what "bullet" means—and the argument can be re-applied to virtually every English word.

So the proposed strategy is impossible. To save the Betan from correction by contingent clues we must try—as with logical clues—to credit him with matching semantic errors; and we have seen what that leads to. This result, like the one in Sect. VIII, is not peculiar to “between.” Other pairs of spatial words certainly yield the same result, and I conjecture that the story would run in essentially the same way for any meaning-switch involving a pair of spatial expressions, just so long as it was not logically equivalent to the “left”/“right” switch. I shall give some evidence for this in Sect. XIII.

XI. CONTINGENT CLUES: THE ALPHAN

Here, perhaps, are some contingent clues for the Alphan:

- (a) “Most clock-hands move downward while to the right of center and upward while to the left of center.”
- (b) “I, like most people, am stronger in my right hand than in my left.”

These, on the Alphan’s interpretations of them, may conflict with generalizations which he knows to hold true in Alpha, i.e., in that part of the world of which he has knowledge not gained through reading English. Can he disarm them by supposing that England differs from Alpha in the relevant respects? Yes, he can. This strategy is open to him as it was not to the Betan, for reasons which constitute the second big contrast between enantiomorphism and betweenness.

First, there are fewer contingent clues to the Alphan’s error than to the Betan’s. For every true generalization that becomes false under the “left”/“right” switch there are hundreds that become false under the transformation of “ x is between y and z ” into “ y is between x and z .”

Secondly, the beliefs about England which the Alphan must initially adopt under his clue-canceling strategy will include only such items as that the English are mostly lefthanded, that their hearts are on the right, that their clocks run counter-clockwise. None of these will ramify, demanding more and more suppositions about matters not directly concerning right and left.

Thirdly, each generalization which is challenged by a contingent clue to the Alphan’s mistake concerns a relatively limited class of things. Where the Betan has to suppose the falsity (in England) of laws of elementary impact-mechanics which govern

the behavior of all middle-sized objects, the Alphan has only to suppose the falsity (in England) of certain generalizations about (i) classes of artefacts and other upshots of human decisions and conventions, and (ii) certain biological species. With one exception from sub-atomic physics, which I shall discuss in Sect. XIV, the only generalizations I know of whose truth-value changes under the “left”/“right” switch are ones which quantify over classes of one of these two kinds.

So the Alphan can easily believe what his clue-canceling strategy requires him to believe. (i) Since the kinds of asymmetry in clock-movements, alphabets, rules of the road, positions of guests of honor, etc., are all matters of social choice, it is likely enough that in England “they order these things differently.” (ii) Nor should the Alphan find it unbelievable that in England the relevant biological generalizations are false; for this is just to suppose that England differs from Alpha in its basic stock of biological material, like the supposition—which would be very believable if our planet weren’t so well explored—that on some Pacific island there are green sparrows and white crows. It would be different if the Alphan had to suppose that England contains animals with the proportions of mice and the bulk of elephants: he would choke on this, because it involves a ratio of leg-thickness to body-weight which goes against certain elementary and basic physical generalizations that hold true in Alpha. But nothing like that is involved in supposing that Englishmen are mostly lefthanded, or in supposing, of a certain species of asymmetrical Alphan snail, that they do not occur in England though their incongruous counterparts do. .

Another point worth noticing about these biological truths that become false under the “left”/“right” switch is that most of them give rather specialized information. The strength of human hands and the placing of human hearts are exceptions to this; but I can think of no other generalizations of this kind which would be known to everyone who led a full, normal, observant, intellectually active life. This is in striking contrast with the ones the Betan has to wrestle with. In Sect. XIV I shall revert to this point.

XII. THE AMBIDEXTROUS UNIVERSE

There are endless matters which might seem to give the Alphan contingent clues which he cannot easily cancel by the proposed strategy. For

guidance on these, and for other pleasures, see Martin Gardner's exceptionally fine book *The Ambidextrous Universe*. I shall discuss a few "pseudo-clues" which I have found to be popular, showing that each fails in at least one of the three following ways: it is not a clue, because the generalization involved does not become false under the "left"/"right" switch; it is not a legitimate clue because it breaks the rule forbidding reference to independently known particulars; it is a clue which can easily be canceled by the proposed strategy.

Mechanical phenomena won't correct the Alphan's error, but it is not obvious that this is so. Given a layout of billiard balls on a billiards table, and a choice of two ways (differing only as right and left) of striking the cue ball, the choice may make a big difference to the final positions of the balls. Does not this supply a basis for contingent clues for the Alphan? It does not. If the initial layout is symmetrical, then the result of striking the cue ball one way will be an incongruous counterpart of the result of striking it the other way, and so for the Alphan all will go smoothly. If the initial layout is not symmetrical, then the Alphan—interpreting our description of it according to his semantic error—will begin not with our initial layout but with its incongruous counterpart; and *then* striking the cue ball one way will give him a final position which is an incongruous counterpart of the one we got by striking it the other way; so again he will have no grounds for suspecting error. This example fairly illustrates the situation with regard to the entire range of mechanical phenomena.

Nor is there any guidance for the Alphan in the common run of electrical phenomena. Rules of thumb relating current-flow to direction of magnetic field, etc., will simply lead him to switch "north" and "south" as applied to magnets; and, short of the *recherché* matter to be discussed in Sect. XIV, that switch would not ramify through causal laws or semantic links.

Of two enantiomorphic forms of a certain acid, only one reacts in a certain way with quinine. But that is a fact about the (asymmetric) form of quinine which happens to be the only one biologically available on our planet. Its enantiomorph is chemically and (given the right stock) biologically possible, and it would react in the given way with the other form of the acid in question. Like all other pseudo-clues involving organic molecules, this falls under the heading of generalizations over certain biological species.

As I implied in Sect. IX, the Alphan can get

logical clues from the interrelations of "north," "east," etc., and so we must credit him with a meaning-switch in respect to these too: specifically, he must think that the orientation of any English map can be expressed in English by the pattern

N

E W, suitably rotated. He may arrive at this

S

through reading "As one stands facing north, east is to one's right"; or, more elaborately, through reading how places in England relate to one another, these relations being expressed both in terms of "left" and "right" and in terms of compass-points. In the latter eventuality, he will

N

find that E W works beautifully on the map of

S

England which he gradually builds up. It will of course be a mirror-map of England, but it will give him no trouble unless he gets some independent knowledge of England—e.g., by trying to tour it with the aid of his map.

If he has a correct map of Alpha, can he com-

N

fortably impose E W upon it? He has no right to

S

assume that it belongs on his map at all; but never mind that. If he does try to impose it on his map of Alpha—or on Alpha—will he encounter any positive obstacles which will serve as contingent clues? To do so, he will have to have (1) something

N

dictating how E W should be rotated before being

S

placed on the map of Alpha, and (2) something else casting doubt on that placement. That is, he needs *two* contingent correlates of compass-points—correlates which he is told are valid in England, and which concern matters in respect to which he cannot easily believe that England differs from Alpha. This might be an example:

"North is the direction toward which compass-needles point. East is the direction from which the sun rises."

One point to notice about these double-correlates

N

which are needed if the Alphan's E W is to yield

S

contingent clues is that if they can generate contingent clues at all they can do so without reference to "north" etc., thus:

"As one stands looking in the direction toward which compass needles point, the rising sun is toward one's right."

The vital point, though, is that the Alphan cannot have even *one*, let alone the needed two, of these correlates of "north," "east," etc. That is, he cannot have good reason to think that any such correlates which he knows to obtain in Alpha must also hold good in England.

Compass needles cannot help us to correct the Alphan, because they point south as well as north. That their ends are differently shaped, and how they are shaped, is a matter of convention.

Still, let us concede compass needles in order to get the sunrise to work. If the Alphan is to get a contingent clue from this, he must say: "Surely the compass-direction of the sunrise in England must be the same as in Alpha!" But why should he say this? Not because a particular star shines on a particular rotating planet which contains both England and Alpha. Of the items which the Alphan knows about in ways other than by reading about them in English, he must not identify any one as *the item* to which the Englishman refers as "the sun" or "the earth (Terra)," though he may recognize some as *items of the kind* the Englishman calls "sunlight" or "stars," "ground" or "planets." (I repeat that this niggardliness is not *ad hoc* or arbitrary. If the Alphan can read English statements about "the earth" and "the sun," and identify these with items known to him in other ways, then he might as well read about and independently identify the constellation Orion, or the box *A* in Sect. IV above, or his right shoulder. From the point of view of the Kantian Hypothesis, any such use of an independently known particular is on a par with our touching the Alphan's right shoulder while saying "I am now touching your right shoulder." This does not make the Hypothesis trivial: its rules for the Betan are just as stern, yet *he* is deluged with logical and contingent clues to his semantic error.)

To mention just one more popular pseudo-clue: since the Alphan may not identify a particular planet—let alone its Northern Hemisphere—as the one containing both England and Alpha, he cannot have any contingent clues involving the direction from which the cold winds blow, or the like.

I cannot anticipate and criticize every plausible pseudo-clue to the Alphan's mistake, but my treatment of the ones I have mentioned may help to show how others should be dealt with.

It is time to consider what the Alphan is to make of the samples of English writing he receives. Clearly, he must not encounter English statements about these samples; and indeed if he encounters English statements about English writing in general—statements which become false under the "left"/"right" switch—then he must suppose that his English correspondent eccentrically writes mirror-English, or that his missives come through a censorship office which photographs them and forwards the negatives, or some such nonsense. These are trivial details. What is not trivial is the following question. Suppose that the Alphan is somehow deprived of samples of written English, but is sent—in Morse-code English, say—very full instructions for writing English letters and words and sentences and paragraphs: what will he write if he follows those instructions as he, with his "left"/"right" switch, understands them? The answer is that he will, without any hitch or hesitation, write perfect mirror-English letters and words and sentences and paragraphs. Sceptics should try it for themselves. There are no clues for the Alphan here.

Of course it would have saved trouble if, following Borel (§§33–36), I had at the outset explicitly placed the Alphan and the Betan on a cloud-covered planet at a great distance in an unknown direction from ourselves. This would have had us communicating in Morse-code from the start; it would also have automatically ruled out all the biological, geographical, meteorological, and sociological overlap between Alpha and England, as well as much of the astronomical overlap; and thus it would have reduced the number of tempting pseudo-clues for the Alphan. I did not adopt this course because, although it would have made things easier, it would not have made clear just what sorts of overlap were being excluded or why they were being excluded.

XIII. SOME OTHER SWITCHES

When I first worked on this topic I contrasted left/right with large/small, but was charged with unfairly exploiting the fact that large/small is metrical. So I re-worked the contrast using "between" instead. The latter, like anything else I might use, also has special features; but they have not essentially contributed to the contrasts I have drawn. To get *prima facie* evidence for this, consider how the story would go for certain alternatives to "between."

Had the switch involved "large" and "small" and their grammatical cognates, there would have been such logical clues as:

"My house is smaller than Jones's—indeed it is the same size as Jones's largest room."

This and its like would require matching semantic errors involving "part" and "whole," "inside" and "outside," "contain," "surround," and hundreds more. And they would ramify: for example, "surround" would infect "grasp," "penetrate," "hole," etc. There would also be contingent clues:

"I couldn't see the water because the house between myself and the shore was so large."

"The rock wasn't small enough for a child of ten to lift it."

It is clear that a "large"/"small" switch would be Betan rather than Alphan.

Suppose we had tried a "round"/"square" switch. These words connect through logical clues with "angle," "smooth," "equidistant," "straight," "curve," "circle," "triangle," and so on. And there would also be many contingent clues involving the role in English life of wheels, building bricks, land surveys, tree trunks, and so on. For example,

"Roundabouts are so-called because the path of someone riding on one is *round*. This is because roundabouts are built and operate as follows . . .",

with the blank filled by a correct account of how roundabouts work. Someone who had switched the meanings of "round" and "square" would have to adjust his semantics and/or his English physics in such a way that that account really would explain to him why the path of someone riding on a roundabout is *square*. Another Betan situation.

Perhaps I needn't offer details on "near" and "far," or "inside" and "outside," or "toward" and "away from." These will very quickly connect, causally and semantically, with "large" and "small"—and we have seen where that switch leads.

What about "head"/"feet" and "front"/"back"? Either of these switches would generate a Betan situation, though an uninteresting one—like a switch in the meanings of "teapot" and "breadboard," or of "nose" and "elbow." It could seem interesting only to someone who was still in thrall to the mistaken view, discussed in Sect. IV above, that the human body is essential to enantiomorphism rather than merely the basis for some convenient terminology for it.

The question "What would happen with an 'up'/'down' switch?", though it could be motivated by the same mistake, has more inherent interest. It is in fact hard to decide what a "switch in the meanings of 'up' and 'down'" would be. The word "down(ward)" can be defined as "the direction of normal fall" or "the direction toward the ground" or both; and analogously for "up(ward)." This suggests three possible interpretations of the switch, but I cannot control the details of any one of them. Part of the trouble is that the logical/contingent dichotomy, which has served well enough for the other switches, lets us down here. "The direction of normal fall" connects with "the direction toward the ground" at least to this extent: if those two directions were different we should have no objects left except ones that were fixed to the ground. Is that a merely contingent matter?

Still, without knowing just what an "up"/"down" switch would be, I think I can show that it would have to be Betan rather than Alphan.

Suppose the contrary. Suppose there is a Gamman who understands English except that he has switched "up"/"down" and certain related pairs such as "above"/"below." He reads many statements about England, interprets them according to his matching set of semantic errors, and believes them. If he is to be analogous to the Alphan, the Gamman's false beliefs about England must not be so far-reaching as to conflict with the postulate that he mainly understands English. So we must suppose that if he came to England with expectations based on what he had read, he would find it fairly much as he had expected, and yet unlike his expectations in some systematic respect.

What respect? The fact that we don't bounce around on our heads? The fact that unimpeded objects don't tend to shoot up into the sky? Anything along the lines of either of those mistakes would obviously generate a Betan state of affairs, not an Alphan one. The only other suggestion I have heard or can think of is this: "He expects unimpeded objects to shoot upward toward the ground—i.e., he has false beliefs about the direction of object-fall in England *and* about where the English ground is in relation to the English sky." But those two "false beliefs" cancel out. The Gamman, in this version of him, would get no surprises if he arrived in England; which is just to say that his English reading has not been infected by any semantic error.

What about a switch in the meanings of "before" and "after"? Nothing in my paper requires an

answer to this question; and that is just as well, for even a sketchy answer—which is all I know how to give—would take up far too much space.

XIV. THE CONSERVATION OF PARITY

The principle of the Conservation of Parity says in effect that if in any general truth of physics we substitute "right" for "left" and vice versa, the result will also be a truth of physics. This principle, though long accepted, is now thought to be false: work which began in Princeton in 1956 has satisfied physicists that there is a basic physical law which becomes false under the "left"/"right" switch.

The point could be put as follows. Suppose that we have two experimental set-ups with initial states I_1 and I_2 and resultant states (arising from the initial ones in ways that can be wholly explained by basic physical laws) R_1 and R_2 . The Parity principle implies that if I_1 is an enantiomorph of I_2 then R_1 is an enantiomorph of R_2 ; and it now turns out that this is sometimes false.

Our Alphan can now be exposed to contingent clues with far more force than any so far mentioned. Let us send him a description of a Parity-refuting experiment, with initial state I_1 and resultant state R_2 , and suppose that he tries to reproduce the experiment in Alpha. He will in fact start not with I_1 but with its enantiomorph I_2 , and he will end up with a resultant state R_1 which is *not* an enantiomorph of R_2 . So the resultant state won't be the one the Englishman predicted, nor will it be the one that the Alphan, with his semantic error, thinks the Englishman predicted. This should lead the Alphan to suspect that he has misunderstood the English description of the experiment, and if he perseveringly tests that suspicion he will learn what the misunderstanding was.

If we are to prevent this, we must try one of two courses.

(i) We may try to credit the Alphan with further semantic errors such that he will understand the English description of R_1 as an accurate description of R_2 . But this takes us back into Betan territory; for these semantic errors—concerning words which are not meaning-linked with "right" and "left"—would ramify into the rest of the language. I believe the Alphan would in fact have to switch the meanings of "more" and "fewer" or some equivalent pair of terms (see Gardner and Frisch), and that switch would obviously lead the whole Alphan story to collapse.

(ii) We may present the Alphan as supposing that if he had performed his experiment in England it would have come out differently. But if he thus distinguishes sub-atomic particles into "the ones we have here in Alpha" and "the ones they have there in England," he will be cutting a very poor figure. Anything which seems to be a fundamental physical law may turn out to be a relatively local accident, but a scientist in his right mind would not accept such a conclusion if he could rationally avoid it. The Alphan has an alternative staring him in the face, namely that he has switched the meanings of "right" and "left" and therefore constructed I_2 instead of I_1 .

At long last, we have got him. (I here suppress more recent developments and discoveries in the physics of this matter, as they lie beyond the scope of my present concerns. For details, see Gardner, *op. cit.*)

This certainly refutes the Kantian Hypothesis as I formulated it: we can now *tell* the Alphan which is which as between right and left. But then we could have told him anyway, using "port" and "starboard." The fact is that the Kantian Hypothesis has served less as a sharp-edged proposition whose truth was to be tested than as a guide to the exploration of some contrasts between enantiomorphism and other spatial distinctions. And those contrasts have not been entirely lost.

Despite the failure of the Parity principle, it remains true that the left/right distinction constitutes, so far as meaning-relationships go, a self-contained unit with simple internal relations and no external relations—that is, no ramifications into the rest of the language.

What of the other contrast? Well, the generalization which we are now using to correct the Alphan is not one which he can easily suppose false in England—it is neither sociological nor biological, but is a matter of fundamental physics. Still, there is a contrast between it and the kinds of causal law which were available for correcting the Betan. Quite recently, two Nobel Prizes were awarded for the discovery of a physical law which does not survive the left/right switch, and so knowledge of that law is a perfect paradigm of specialized, non-common knowledge. Someone who grasps all the underlying fact and theory will not find the law "easy to suppose false"; but we can lead busy, observant, intelligent lives without having the slightest need to think that the law is true. This situation could change, if our technology came increasingly to depend upon the law in question;

but knowledge of that law and of any asymmetries depending upon it is, and will long continue to be, optional intellectual equipment. So, even with the failure of the Parity principle taken into account, the right/left distinction (by which, always, I mean the distinction between any enantiomorphic pair) still differs enormously from every other spatial distinction: it remains unique in its degree of isolation in the layman's language and the layman's *Weltanschauung*. Is it too ambitious to suggest that these simple facts help to explain physicists' surprise at the failure of the Parity principle?

University of British Columbia

Received June 9, 1969

BIBLIOGRAPHY

- AVERROËS, *Tahafut al-Tahafut* (trans. S. van den Bergh, London, 1954).
- BOREL, Émile *Space and Time* (New York, 1960).
- CAIRD, Edward *A Critical Account of the Philosophy of Kant* (Glasgow, 1877).
- FRISCH, O. R. "Parity not Conserved: a New Twist to Physics?", *Universities Quarterly*, vol. 11, (1957).
- FRITSCH, Vilma *Left and Right in Science and Life* (London, 1968).
- GARDNER, Martin *The Ambidextrous Universe*, revised edition (New York, 1969).
- GHAZALI, Abu Hamid Muhammad *Tahafut al-Falasifa* (trans. S. A. Kamali, Lahore, 1958).
- JAMMER, Max *Concepts of Space* (Cambridge, Mass., 1954).
- KANT, Immanuel "Von dem ersten Grunde des Unterschiedes der Gegenden im Raume" (1768), in *Gesammelte Werke* (Akademie Ausgabe), vol. 2. Translated as "On the First Ground of the Distinction of Regions in Space" in John Handyside (trans.), *Kant's Inaugural Dissertation and Early Writings on Space* (Chicago, 1929).
- KANT, Immanuel, Inaugural lecture: "De mundi sensibilis et intelligibilis forma atque principiis" (1770), in *Gesammelte Werke* (Akademie Ausgabe), vol. 1. Translated as "Dissertation on the Form and Principles of the Sensible and Intelligible World" in Handyside (see preceding entry).
- KANT, Immanuel *Prolegomena to any Future Metaphysic that will be able to present itself as a Science* (trans. P. G. Lucas, Manchester, 1953). Reprinted in Smart.
- KEMP SMITH, Norman *A Commentary to Kant's Critique of Pure Reason* (New York, 1962).
- LEIBNIZ, G. W. *The Leibniz-Clarke Correspondence* (ed. H. G. Alexander, Manchester, 1956).
- MAYO, Bernard "The Incongruity of Counterparts," *Philosophy of Science*, vol. 25 (1958), pp. 109-115. A reply to Pears.
- PEARS, D. F. "The Incongruity of Counterparts," *Mind* n.s. vol. 61 (1952), pp. 78-81.
- REICHENBACH, Hans *Philosophy of Space and Time* (New York, 1958).
- REMANT, Peter "Incongruent Counterparts and Absolute Space," *Mind* n.s. vol. 72 (1963), pp. 393-399.
- RUSSELL, Bertrand *Essay on the Foundations of Geometry* (Cambridge, 1897).
- SCOTT-TAGGART, M. J. "Recent Work on the Philosophy of Kant," *American Philosophical Quarterly* vol. 3 (1966), especially pp. 178-180.
- SMART, J. J. C. (ed.) *Problems of Space and Time* (New York and London, 1964).
- WEYL, Hermann *Philosophy of Mathematics and Natural Science* (Princeton, 1949). See also the same author's *Symmetry* (Princeton, 1952), especially pp. 16-38.
- WITTGENSTEIN, Ludwig *Tractatus Logico-Philosophicus* (New York, 1961).

² A version of this paper was read to several philosophy departments in 1966 and 1967, and was much helped by the comments and criticisms of many people to whom, though I cannot separately name them, I here express my sincere thanks. In other ways the paper has benefited from the generous collaboration of Lewis White Beck, D. G. Brown, Martin Gardner, James G. Hopkins, and Douglas F. Wallace, to all of whom I am deeply grateful.

II. WITTGENSTEIN, PRIVILEGED ACCESS, AND INCOMMUNICABILITY

RICHARD RORTY

I. INTRODUCTION

IN this paper, I wish to argue for the following theses:

- (A) None of the arguments about the possibility of a private language or about the privacy of sensations and thoughts which Wittgenstein advances in the *Philosophical Investigations* provide good reason for doubting
 - (a) that words like "toothache" and "pain" are the names (in a nontrivial sense) of sensations which people sometimes experience, or
 - (b) that when I assert truly "I have a toothache" or "I am in pain," I am describing the state of my consciousness, or
 - (c) that when I assert of another person "He has a toothache" or "He is in pain" I claim that he is experiencing the same sort of sensation that I do when I have a toothache or am in pain
- (B) None of these arguments give good reasons for rejecting as senseless the claim that "sensations are private"
- (C) None of these arguments give good reasons for rejecting as senseless the claim that "I know that I am in pain because I feel it."

I have drafted these theses with an eye to recent discussions of Wittgenstein's views about the privacy of sensations, and in the belief that certain confusions committed by Wittgenstein or his interpreters—notably between "privacy" in the sense of "susceptibility to privileged access" and in the sense of "incommunicability"—have led sympathetic commentators to attribute unnecessarily paradoxical views to him, and hostile critics to attack him by attacking these paradoxes.

Wittgenstein's central insights have thus, I believe, been obscured. Accordingly, this article will consist of commentary on commentaries on Wittgenstein. By reviewing the literature on the topic, I hope to make possible a fresh look at some of Wittgenstein's central themes.

The wording of clauses (a), (b), and (c) in (A) above are taken from George Pitcher,¹ who argues that these three clauses express the view which Wittgenstein believed his arguments to have overthrown. (Pitcher does not believe that these arguments *do* overthrow this view, but many of Wittgenstein's readers have thought they did.) John Cook,² in contesting Pitcher's account of what Wittgenstein thought he had shown, argues that Wittgenstein's aim was rather to show the senselessness of the assertions quoted in (B) and (C). In Sect. (II) I argue for (A) by rebutting arguments which Pitcher attributes to Wittgenstein. In Sect. (III) I argue for (B) and (C) by rebutting those attributed to him by Cook. (I shall not attempt to settle the question of whether Wittgenstein actually advanced either set of arguments.)

II. PITCHER'S INTERPRETATION

Pitcher calls the view expressed by the three clauses cited under (A) above "View V" and claims that they—and, in particular, (b)—entail that "I can never know for certain whether . . . another person is in pain or not—for I cannot feel another person's pain."³ He regards this as a *reductio*, but does not argue for the entailment. Bating the question of whether the conclusion is absurd, it would seem that the entailment would only hold if

I know for certain that Jones is in pain
entailed

I feel Jones's pain.

¹ *The Philosophy of Wittgenstein* (Englewood Cliffs, New Jersey, 1964), p. 285.

² "Wittgenstein on Privacy," *Philosophical Review*, Vol. 74 (1965), pp. 281–314. Reprinted in *Wittgenstein: The Philosophical Investigations: A Collection of Critical Essays*, ed. by George Pitcher (New York, 1966), pp. 286–323.

³ Pitcher, p. 285.

But one would accept this latter entailment only if one held

- (1) If Jones is certain that he is in pain, and if I am certain that he is in pain, we must both have the same grounds for certainty

and there seems no plausibility to this view.

Again, Pitcher claims that V entails that "I cannot conceive that another person feels the same sensation that I do when I feel a pain."⁴ He supports this by saying that

There are in fact no specifiable conditions under which I could determine that another person feels the same sensation I do: to do that I would have to be able to feel his pain (see *PI*, sect. 253); and that is impossible. He can, of course, describe his pain to me as "sharp," "dull," "severe," and so on, but this is no help whatever; for I have no way of telling what corresponds to these adjectives in this case. The adjectives are here being used analogically. . . . Since there is no way of specifying how the truth of the assertion "He feels the same sensation I do" could possibly be determined, the assertion is unintelligible.⁵

In this passage, and elsewhere, Pitcher is assuming that view V entails

- (2) "Pain" is the name of a kind of private sensation

and that this in turn entails

- (3) I only learn what pain is from my own case which in turn entails

- (4) I can only apply the word "pain" to my own sensations.

But it is not clear why (2) entails (3). Here we encounter the problem of the meaning of "privacy," and it will therefore be helpful to recall some distinctions, drawn most explicitly by Ayer, between possible meanings of this term.⁶ Ayer notes that we might say

- (P₁) Things are private to a given person if their existence, or one or more of their qualities, could be known by him but not conceivably by anybody else.

or

- (P₂) Things are private to a given person if there is at least one way in which he can detect either their existence, or one or more of their qualities, but others cannot.

or

- (P₃) Things are private to a given person if his authority about either propositions like "There is an . . ." or "There is an . . . which has a certain quality" cannot be overridden.

or

- (P₄) Things are private to a given person if he has something which he can either not share with another person, or which he can share only partially (in the sense that he can have the same *kind* of thing as another person, but not the numerically identical thing).

Now View V does not commit us to (2) even if (2) is interpreted in the weak sense of (P₄). (When so interpreted, we should notice, (3) does not follow.) As it stands, View V says nothing about epistemological questions, and thus does not commit us to the privacy of sensations in any other sense. However, Pitcher is, reasonably enough, assuming that one who holds V will also hold

I know about my pains by feeling them
and

Nobody else can feel my pains, nor can I feel anybody else's.

These propositions would commit the holder of V to interpreting (2) in the sense of P₂ as well as that of P₄. But even this would not allow the inference to (3). For we might well say that I can learn what pain is (and/or what "pain" means) in all sorts of complicated ways, and could learn this even if I never happened to have had a pain. Even if we granted that we could not learn what pain was unless we had received appropriate linguistic training concurrently with being in pain—thus granting (3) though not the entailment of (3) by (2)—one would still not be in a position to go from (3) to (4) without the help of some such premiss as

- (5) If I learn the meaning of a word by ostension (i.e., if ostensive definition is a necessary part of the process of learning the meaning of the word), then I can meaningfully apply that word only to objects ostensible to me in the same way in which the original cases (the ones used in training me) were ostensible.

⁴ *Ibid.*, p. 288.

⁵ *Ibid.*, p. 288.

⁶ Cf. *The Concept of a Person* (New York, 1963), p. 79. In stating these possible meanings, I have touched up Ayer's statement of them by speaking of "the existence of, or some quality of" an object, rather than simply of "the existence of" the object.

But nothing in View V seems to support such a *prima facie* implausible premiss. Why should one who holds V not answer Wittgenstein by saying that the criteria for the truth of "He feels the same sensation I do" are the standard behavioral and environmental ones? Pitcher implicitly grants the possibility of this reply when he says that "The foregoing considerations [various arguments, including the one now under discussion] do not show that 'pain' is not the name of a private sensation. They show only that some outward manifestations of pain are required for the teaching and learning of the use of the word 'pain'."⁷ Pitcher says the admission of this latter point constitutes a "modification" of V, but it is not clear that it is. He so describes it, I believe, because he implicitly assumes that one who upholds V is also committed to

If 'N' is the (common) name of things of a certain kind, then (for at least some kinds) nothing save awareness of such things, plus correlated utterances of 'N' by those who know the language which contains 'N', are required to learn what 'N' means.

This assumption is one which I think Wittgenstein *does* have a good argument against. Here, however, I want simply to note that it is logically independent of V. First, however, for the sake of completeness, let us look at a final argument which, according to Pitcher, Wittgenstein advances against the "unmodified" form of V.

This argument rests on the premiss that "The expression of doubt has no place in the language-game" (which we play with utterances like "I am in pain now").⁸ Pitcher presents Wittgenstein as saying that if V were correct then "when a sensation appears before a person's mind, he must identify that item (as, say, a pain rather than itch or an ache or a twinge). In that case, the possibility arises at once that he might make a mistake in his identification. He might always misidentify it, and hence it must always be possible for him to wonder whether he has done so or not."⁹ But the quoted premiss shows that he cannot so wonder. Pitcher is here suggesting that Wittgenstein infers from

(I) I can identify a certain particular

to

(II) It is in principle possible for me (given present practices, and barring misuse of language, or failure to master the language) to make a mistake in identifying

that particular (a mistake which is not a result of ignorance of the language—of, e.g., the meanings of terms like "pain," "sensation," "stabbing," etc.)

to

(III) I have criteria for identifying that particular.

But, defenders of V may reply, why does not the case of mental particulars simply show that the inference from (I) to (II) is fallacious? Why should all cases of identification be cases of corrigibility?

When we have to choose between an analysis of a concept like "identification of a particular" (e.g., one which would license the inference from (I) to (II)) and a piece of conventional wisdom such as View V, what is our criterion? It is not too much to say that certain interpreters of Wittgenstein have argued as follows: "We can make mistakes when we identify, but not when we express feelings. There are no mistaken pain-reports. Therefore pain-reports are expressions of feeling, rather than identifications of particulars." So put, the argument is to say the least of it, an obvious *petitio*. To get a good argument we need to argue that no analysis of "identifying a particular" which does not permit the inference from (I) to (II) will be adequate. But I do not think that any interpreter of Wittgenstein has offered such an argument. Rather, they have assumed that any adequate analysis of "identifying a particular" will include the notions of "observing the particular (or its effects)," "noting whether the particular satisfies certain criteria," etc. Since they note, rightly, that the Cartesian tradition in philosophy has built up a host of philosophical perplexities out of the assumption that we possess an inner eye which inspects (and, so to speak an "inner mind" which forms judgments about) mental particulars, they assume that the fastest way to overthrow this tradition is to abandon the notion that when we report, e.g., pains or thoughts we are identifying particulars. It is indeed a fast way, but it is certainly not the only way.

More will be heard of the fact that we cannot have doubts about whether we are in pain when we come to Cook's interpretation of Wittgenstein. We may turn now, however, to Pitcher's account of Wittgenstein's arguments against the "modified" version of V. On this account, "the ways in which

⁷ Pitcher, p. 292.

⁸ Wittgenstein, *Philosophical Investigations*, Pt. I, sect. 288.

⁹ Pitcher, p. 290.

the names of public objects and qualities denote their objects cannot be even remotely like the ways in which 'pain' denotes a sensation," and this is seen by noting that

I can do practically none of the things [with pains] I can do with physical objects or colors or shapes, i.e., with publicly observable things, and so the modes of behavior in which alone the connection between the name of something public and the thing it names is made are not available in the case of "pain."¹⁰

The defender of V is then envisaged by Pitcher as replying that the process of learning by ostensive definition which we undergo while coming to use "tree" or "red" correctly may, in fact, be paralleled by a similar process—namely, the one which the celebrated "private diarist" conducts when he decides to call a given sensation 'E'. Having taken this line, he is now a patsy for the familiar objection that "the concept of correctness and incorrectness does not apply" to 'E'.¹¹

Here we need merely ask: is this appeal to a private diary the only rejoinder which one who holds V can make to the argument which Pitcher attributes to Wittgenstein? The difficulty in answering this question is a result of the vagueness of the phrase "the way in which names denote. . . ." How much like the way in which "tree" denotes trees does the relation between "pain" and pains have to be? What parameters of similarity and difference are appropriate here? Does the fact that we can do "practically none of the things" with trees and "tree" that we do with pains and "pain" count as showing that what Wittgenstein calls "the model of 'object and name'" is out of place here? May it not be that the model can be kept, as long as we are on our guard not to assume that, because the same model is used for "tree" and for "pain," everything that holds for the former holds for the latter? By putting the claim that "we can keep a private diary" in the mouth of the holder of V, one seems to assume that

- (6) We cannot come to know what the referent is of an observation-term (i.e., a referring expression which very frequently occurs in non-inferential reports—reports which people make without having gone through a conscious process of inference) except by ostention.

Since "pain," under this definition, obviously is an observation term, the truth of (6) would commit those who hold that "'pain' is the name of private sensation" to something like the "private diary" picture of how we learn the names of mental particulars. But, given Wittgenstein's critique of the whole notion of "ostensive definition," and the doubt this critique casts on (6), cannot we simultaneously repudiate (6) and maintain V? Once again we need to ask whether we cannot simply appropriate Wittgenstein's theories about language and the learning of language without thereby giving up V. As in the case of choosing an analysis of "identifying a particular," there may be a perfectly good analysis of "denoting" which covers both "tree" and "pain." At least we may say that unless it is shown that such an analysis cannot be given, and that any adequate analysis will entail (6), the "private diary" argument does not show that "pain" may not be the name of a private sensation, where "private" has the senses of (P₂), (P₃), or (P₄).

In my examination of the arguments so far, it may seem that I have perversely evaded the point. It seems clear that, despite the gaps in the arguments which I have tried to point out, there is *something* to these arguments, and that our treatment so far has not got at it. I think this is so, and I think that many threads can be knotted together by focusing on the central premiss of the next (and last) argument which Pitcher presents—his interpretation of Wittgenstein's "beetle-in-the-box" argument. This premiss is:

Everyone acknowledges that sensations are private, that no one can experience another person's sensations, so that the special felt quality of each person's sensations is known to him alone and to no other.¹²

What Pitcher does in this sentence is to move from privacy in the sense of (P₄) to privacy in the sense of (P₁)—from

- (7) Sensations cannot be shared

to

- (8) We cannot communicate certain qualities (the "special felt qualities") of our sensations to others.

Given (8), Pitcher can continue

Thus, when you are in pain, I do not know, cannot know, the character of your sensation—whether, for

¹⁰ *Ibid.*, pp. 293–294.

¹¹ *Ibid.*, p. 297.

¹² *Ibid.*, p. 297.

example, it is exactly like what I might feel if my hand were wounded as yours is now, or whether it is something altogether different.

But, restricting ourselves to V and its consequences, this line of argument is no more convincing than the sceptical suggestion that you and I, when we inspect a patch of red or a beetle, are having wildly different experiences or seeing wildly different things, even though we say all the same things. If we both say the same things about our sensations, the patch, or the beetle, then it is not clear what it means to say that there are certain qualities of sensations, patches, or beetles which are "incommunicable." We may wish to squash scepticism about our knowledge of beetles by noting that a difference which is not reflected in a possible difference in what we say is, in some sense, not a real difference. But one may happily grant this point and yet insist that the fact that pains cannot be shared does nothing to show that the relation between "pain" and pains differs from the relation between "red" and red patches or "beetle" and beetles. If one refuses to take scepticism about patches or beetles seriously—on the ground that the sceptic's suggestion that it might be, so to speak, sheer chance that we all agree in what we say about them does not make sense—then one should not take scepticism about pains seriously either, and thus one should not grant (8). Nor does it help to say that it is only the false view of pains as private sensations which licenses the scepticism of (8). For one who holds view V may be committed only to sensations as private in the sense of (P₄) (and possibly (P₃) and (P₂) as well), not to their privacy in the sense of (P₁), and thus not to (8).

Given this point, we can account for our sense that there is *something* to the various arguments which Pitcher attributes to Wittgenstein by noting that if (8) is accepted—or, more generally, if privacy in the sense of (P₂), (P₃), or (P₄) is taken to entail privacy in the sense of (P₁)—then a number of dubitable premisses which we have isolated above—e.g., (1), (3), (5), and (6)—look a good deal more plausible than they do in isolation. They look still more plausible if we adopt two more principles, viz.,

- (9) I know what a given sort of sensation is only because I know about certain incommunicable "special felt qualities" which are characteristic of certain sensations I myself have.

and

- (10) I do not know whether it would be appropriate to apply the name of a given sort of sensation—e.g., "pain"—unless I know whether I am talking about something which has certain incommunicable "special felt properties."

I think that the plausibility of Wittgenstein's arguments, as they have been presented by Pitcher, arise from his tacitly treating (8), (9), and (10) as an intrinsic part of view V. This treatment is by no means disingenuous, for it is the case that *philosophers* who held V had frequently held (8), (9), and (10) as well. But it is nevertheless important both to distinguish between what philosophers have tacitly accepted and what common sense would say, and to recognize that these various theses are at least a few steps farther away from common sense than the three assertions which express Pitcher's "official" version of V.

I want now to finish the job of showing that, in Pitcher's account of Wittgenstein's arguments, *everything* turns on the truth of the additional premisses (8), (9), and (10). So I now come back to the argument which Pitcher constructs on the basis of his premiss about "special felt qualities." Pitcher quotes the celebrated "beetle-in-the-box" passage and says

The analogy with pain is perfectly clear. If "pain" is supposed to denote a somewhat (including a nothing) which each person can observe only in his own case, then the somewhat "cancels out"; and if the sole function of the word "pain" is to denote it, the word is at once deprived of any use.¹⁸

There are various objections which might be made here—notably about the parenthetical clause "including a nothing" and about the attribution to holders of V of the claim that this is the "sole function" of "pain." But let us bate these issues and consider simply what is being said when we say that "the somewhat 'cancels out'." The original passage in Wittgenstein, after describing people's privileged access to the contents of their boxes, says

But suppose the word "beetle" had a use in these people's language?—If so, it would not be used as the name of a thing. The thing in the box has no place in the language-game at all; not even as a *something*: for the box might even be empty.—No, one can "divide through" by the thing in the box; it cancels out, whatever it is.

¹⁸ *Ibid.*, p. 298.

That is to say, if we construe the grammar of the expression of sensation on the model of "object and designation" the object drops out of consideration as irrelevant.¹⁴

Pitcher says that Wittgenstein is here

only denying a particular thesis about language, namely that the word "pain" names or designates this something that the person feels, in a way which is even remotely like the way that words for publicly observable things name or designate them. In the language-games we play with words like "tree" and "red," trees and redness (red things) play some part, and it is in these games that the connection between the name and the thing named is established. But in the numerous language-games we play with the word "pain," private sensations play no part, and so "pain" cannot denote them in anything like the way that "tree," for example, denotes that kind of object. What does play a part in pain language-games is pain behavior . . . and pain-comforting behavior . . . in short, the external circumstances in which the word "pain" is used.¹⁵

Pitcher thus identifies "is canceled out" with "plays no part in the language game." Two points may be made about this passage. First, in saying that "private sensations play no part" Wittgenstein is either committing a *petitio* or else using "private sensations" as an abbreviation for "those special felt qualities of private sensations which are not communicable in language." In the latter case, he is perfectly justified in his claim, but then we must realize that it is not self-evident that there are such qualities, and that even if there were it might be that private sensations had plenty of *other* qualities which *were* communicable. Secondly, consider the following parodies:

But in the numerous language-games which we play with "tree," the tree-in-itself plays no part, and so "tree" cannot denote in anything like the way that words which refer to directly experienced entities (e.g., green sense-data) denote. What does play a part in tree language-games is sense-data, and certain other mental entities (intentions, judgments, volitions, desires, etc.)—in short, the directly apprehended objects of awareness which are before our consciousness in situations in which we learn how to use "tree."

But in the numerous language-games which we play with "neutron," the neutrons themselves play no part,

and so "neutron" cannot denote in anything like the way that words which refer to directly experienced entities (e.g., cloud-chamber tracks) denote. What does play a part in neutron language games are pointer-readings, unexpected results in mathematical calculations. . . .

I take these parodies to show if we use the notion of playing no part in the language-game to explicate "canceled out," it will be much too easy to cancel things out.¹⁶ The trouble is that it is never very hard to describe the process of learning and using a given term—one which, *prima facie*, seems to be used to denote particulars—in a language which makes no reference to these particulars. All that one has to do is to use only terms which occur in criteria for applying the term in question. One will thus be able to argue that the particulars putatively denoted "play no part" in the language-game played with the putatively denoting term, and thus "cancel out."

I have developed the second point to emphasize the importance of the first, and to show that, on Pitcher's interpretation, the "beetle-in-the-box" argument reduces to the argument that what is *incommunicable* can play no part in a language-game. This argument may be expressed as follows:

- (a) Suppose that *Q* is an incommunicable quality of a particular—a quality which cannot be characterized in language
- (b) Let *G** be the (indefinitely large) set of rules which govern the correct application of a (putatively) referring expression '*P*'
- (c) Suppose that we cannot tell whether '*P*' has been correctly applied unless we know whether it is being applied to something that is *Q*
- (d) Then, by (c) *G** must contain some rule of the form (R) "Call it a *P* only if you have good reason to think that it is *Q*"
- (e) But then *G** is inexpressible in language, for, by (a), no expression in any language characterizes *Q*, and (R) is therefore inexpressible.
- (f) But the notion of there being rules for using a language which cannot themselves be expressed in any language is incoherent.

¹⁴ *Philosophical Investigations*, Pt. I, sect. 293.

¹⁵ Pitcher, pp. 298–299.

¹⁶ The fact that such parodies can be constructed has suggested to such writers as Strawson, and Chihara and Fodor, that Wittgenstein is taking for granted the sort of operationalism which inspired Berkeley, the Absolute Idealists, and positivistic phenomenologists. I think that he did indeed do this, confusing the kernel of truth in the verificationist theory of *meaningfulness* with a stronger, false, theory about *meaning*.

- (g) Therefore (a) and (c) are mutually incompatible hypotheses.

If one grants (f), then I think that this argument is sound.¹⁷ But we now see that, the argument—in (c)—presupposes (10)—*viz.*, that the incommunicable “special felt” qualities of my private sensations are of the *essence* of those sensations. The reason why Wittgenstein (according to Pitcher) thinks that private sensations cancel out but that trees and neutrons do not is that the former, but not the latter, are “beyond the reach of language.” If I am right in suggesting that the denials of (8)–(10) are absolutely central to Wittgenstein’s thought, then it is easy to see why Wittgenstein should have attached so much importance to denying that the relation between pains and “pain” can be construed “on the model of ‘object and designation’.” If (8)–(10) were true, then we would have to grant the existence of knowledge unmediated by language—the central Cartesian fallacy. For we would have to grant that words in a public language could be learned and used only by calling upon knowledge of (as yet) incommunicable qualities—the usual Cartesian-Lockean picture of language-learning.

This exegesis of the final argument which Pitcher attributes to Wittgenstein may be confirmed by noting one more remark which he makes:

If, after you say to someone “I am in pain,” he sympathizes with you, comforts you, does what he can to help you, then the word “pain” has done its work—and it was not used to tell him the nature of what you had before your consciousness, because that cannot be told.¹⁸

The first clause here is unexceptionable, but the second clause (“and it was not used . . .”) rests on nothing except the claim that “that cannot be told” and the vague suggestion that if an utterance does one job, it cannot do two. If we discount the latter suggestion (as I think we should) then we are left with the claim that what is before my consciousness “cannot be told.” This claim rests *entirely* on (8) and (10).

III. COOK’S INTERPRETATION

In the article referred to earlier, Cook criticizes Pitcher for attributing to Wittgenstein the view that

“I cannot ‘determine that another person feels the same sensation I do’.”¹⁹ Cook argues that Wittgenstein did not commit himself to (8)—on a construal of (8) according to which the “special felt qualities” are objects of knowledge—nor to (9) and (10). But his argument depends on attributing certain other dubious theses to Wittgenstein, notably the following:

There is no criterion of numerical (as opposed to generic) identity of sensations

The notion that “sensations are private objects” is senseless

It is senseless to say that “I know that I am in pain because I feel it”

Since we have presupposed the contradictories of all three of these theses in our discussion of Pitcher’s interpretation, we need now to inquire whether Wittgenstein provides good arguments in favor of any of them.

Although Cook’s criticism of Pitcher comes at the very end of his discussion, it will be convenient to begin with an analysis of this criticism to connect the remarks I have already made about Pitcher’s version of Wittgenstein with the criticisms I wish to make of Cook’s. Cook says that Pitcher thinks that Wittgenstein is criticizing a common-sensical view (V), whereas he is in fact only criticizing a “philosopher’s picture,” Cook notes that Pitcher does not quote the last sentence of Wittgenstein’s “beetle-in-the-box” passage, *viz.*, “That is to say: if we construe the grammar of the expression of sensation on the model of ‘object and name’ the object drops out of consideration as irrelevant.” He continues

The word “if” here is crucial, for it is not Wittgenstein’s view but the one he opposes that construes the grammar of the expression of sensation on the model of “object and name,” and therefore it is not Wittgenstein, as Pitcher thinks, who is committed to the paradoxical consequence that in the use of the word “pain,” for example, the sensation drops out as irrelevant. The point of the passage, then, is quite the opposite of what Pitcher supposes. Rather than showing that sensations cannot have names, it shows that since the view that sensations are private allows sensations to have “no place in the language game” and thereby makes it impossible to give any account of the actual (that is, the “public”) use of sensation words, we

¹⁷ I think that (f) is true, and that Wittgenstein thought so too. But I do not think he had any convincing arguments in its favor, nor do I.

¹⁸ Pitcher, p. 299.

¹⁹ Cook, p. 513 (322). The number given in parentheses in references to Cook’s article is the page number in Pitcher’s anthology (see footnote 2 above).

must, if we are to give an account of that language game, reject the view that sensations are private.²⁰

Roughly speaking, Pitcher takes Wittgenstein to say that since sensations are private, they can't have names, whereas Cook takes him to say that since sensations have names, they can't be private. Both, then, seem to agree that Wittgenstein needs as a premiss

Private objects can't have names.

I wish to argue that Wittgenstein has given no good reasons for holding this view. In examining it, we are once again led to ask for an explanation of "private," but Cook is no help here. He thinks that it is pointless to criticize Wittgenstein for not having explained more clearly what he meant by "private language" because

the idea under investigation turns out to be irremediably confused and hence can be only suggested, not clearly explained. Moreover, the philosophical idea of a private language is confused not merely in that it supposes a mistaken notion of language (or meaning) but in its very notion of the privacy of sensations.²¹

Cook here seems to be saying that the very notion of a "private object" is too confused to be explained. In the light of Ayer's vigorous efforts to formulate explanations of various senses of this notion, this is a strong claim. Cook backs it up by claiming that only a false presupposition that sensations can be numerically identical permits Ayer to formulate his explanations. Thus, according to Cook, when Ayer says that in the sense in which two people can have the same pain or the same thought we do *not* have "numerical identity" in mind, he is presupposing the view that sometimes we *do* have numerical identity in mind, and this is false. Cook thinks that we must answer "no" to the question "is there, then a familiar use of sensation words with a criterion of identity that is reflected in 'But surely another person can't have this pain?'" His argument for this negative answer comes out best in the following passage:

Thus, if a mother has described one of her children's tantrums, someone else might remark that her child had had a tantrum "exactly like that" . . . "Exactly like" is used here in contrast, not with "same," but with "rather like," "rather different," and so forth. That is, it would not be asked: "Do you suppose they may have had the same one and not just two exactly

alike?" This kind of identification question has no place in the grammar of "tantrum," and so neither do its two answers: "Yes, they did have the same one" and "No, they did not have the same one, only two exactly alike." Now this same point holds for the grammar of "toothache" . . . That is, it would not make sense to say, as if in answer to that question, *either* "They had the same toothache" or "They did not have the same toothache."²²

We can agree that such questions "should not be asked," but we can also hold that the reason they are not asked is not that we lack a criterion for numerical identity of toothaches, but simply that it is obvious that the criterion for numerical difference is satisfied. Consider the following definition:

- (D) Two sensations are generically the same if and only if the persons who have them describe them in exactly the same terms, answer questions about them in exactly the same way, etc., and two generically identical sensations are numerically different if and only if they are had by different persons or by the same person at different times.

What is wrong with (D)—simple-minded as it is—if one wants a criterion for the numerical identity of toothaches? Cook is right in suggesting that Ayer would not accept it, since Ayer wishes to leave open the possibility of "co-consciousness," and thus does not want its impossibility built into the language-game. But suppose that, for the moment, we just decide to describe all cases of possible "co-consciousness" (e.g., cases of split personality, cases of interchanged or inter-communicating organs, etc.) in a way compatible with (D). Do we not then have all we need to give sense to the notion of "generically the same, but numerically different, sensations"?

Cook would presumably say that we do not and would insist that

- (I) If a yes-or-no question does not normally occur in extra-philosophical discourse (although all the words used in it do), then the question, and all direct answers to it, are senseless.

Cook regards this principle as superior to the one usually adopted by philosophers who claim that "It is a necessary truth that no two people can have the same toothache"—*viz.*, a principle such as

²⁰ *Ibid.*, p. 312 (322–323).

²¹ *Ibid.*, p. 281 (286–287).

²² *Ibid.*, p. 308 (312).

- (II) If a yes-or-no question does not occur in extra-philosophical discourse because it is never the case that the answer to it is "yes" (or because it is never the case that the answer to it is "no"), then a suitably generalized form of the answer which *would* invariably be given (if the question *were* asked) expresses a necessary truth.

One may sympathize with Cook's doubts about (II), since the notion of "necessary truth" (or "conceptual truth" or "grammatical truth") is indeed very murky. But I think the attempt (which is a major theme of Cook's article) to substitute notions of "confusion arising from mixing up distinct language-games" and of "senselessness" for such notions is a cure which is worse than the disease. One can applaud Cook's remark that "This talk of grammar 'forbidding us' to say something is nothing but the most recent jargon for calling a halt to an analogy whose oddness has begun to dawn on one,"²³ while regretting that Cook himself does not stick to blocking bad analogies, rather than resorting to charges of "senselessness" and invoking (I). To see what is wrong with (I), consider the following argument against the claim that there is a criterion of numerical identity for mountains:

It might be discovered that two mountains—one in Alaska and one in Antarctica—were exactly like each other. They had precisely the same configuration, the rock of which they were made seemed qualitatively identical—and so on for all features within a radius of, say, two air miles from their respective peaks. "Exactly like" is used here in contrast, not with "same," but with "rather like," "rather different," and so forth. That is, it would not be asked "Do you suppose they may be the same mountain?" This kind of identification question has no place in the grammar of "mountain," and so neither do its two answers: "Yes, they are the same" and "No, they are not."

This argument does not show anything except that if we know that two mountains have different locations, we don't ask whether they are the same mountain. It certainly does not show that "The same mountain cannot be in two different places at the same time" is senseless. So *sometimes* (I) is false.

On the other hand, Cook has an example which makes (I) look fairly plausible. This is the noun "build," as in "His build is exactly like his father's." Here indeed the question "Do they have the same build?" cannot be construed in different ways—once in regard to qualitative identity and again in

regard to numerical identity. Further, the answer "No, the builds are different, but they're exactly alike" is "senseless," if any grammatical English sentence is. But this example just forces us to ask: (a) "Are pains more like mountains or like builds?"; (b) since (D) would, *mutatis mutandis*, seem to work for builds as well as for sensations, is not our so-called criterion for numerical identity of sensations susceptible to a *reductio ad absurdum*?

The answer to (a), I think, is that pains are more like mountains simply in the respect that we *do* sometimes say things like "But nobody can have *this* pain" and "But the one in Antarctica can't be *this* mountain" and *don't* say things like "But nobody can have *this* build." (If it were said that nobody but a philosopher would use the first of the quoted sentences, I would not wish to argue a point which seems to me irrelevant. It would be relevant only if philosophers played on the limited—and, to my mind, quite innocent—analogy between pains and mountains created by their odd linguistic habits in order to infer to some *further* analogy.) There is no particular reason, as far as I can see, why we treat pains but not builds as particulars, but then why shouldn't we? The answer to (b) is that we *could* treat builds as particulars, in which case we doubtless *would* use a modification of (D) as a criterion of numerical identity for builds. It just happens that we don't.

In making these points, I have no wish to deny Cook's point that builds and pains are analogous in a way that neither are analogous with coats. Cook says that the "my" in "my coat" is a "possessive of ownership" because the question of whether it's my coat (whether I own it) is not settled by the fact that I've got it. In this (stipulated) sense, the "my's" in "my build" and "my toothache" are, to be sure, not possessives of ownership. But Cook goes from

In cases where "my" is not a possessive of ownership, "is my *X* the same as his *X*?" is not a "genuine identification question"²⁴

to

"Is my *X* the same as his *X*?" is senseless, and so are all direct answers to it,

and I do want to criticize this inference. If a "genuine identification question" means "one that would not be asked by anybody who knows the language" then we still have to ask "Do we not ask it because the answer is so obvious, or because we

²³ *Ibid.*, p. 303 (311).

²⁴ Cf. *ibid.*, p. 296 (303).

can't make sense of it?" In the case of builds, I am inclined to say "because we can't make sense of it," although I can imagine someone making sense of it by creating a new language-game in which "my build" is treated as the name of a particular, rather than as the name of a collection of qualities. In the case of pains, I am inclined to say "because the answer is so obvious." All that Cook has done is to note that there would be no great difficulty in treating "my toothache" as the name of a sharable collection of qualities. He has not shown that we *do* so treat it or that we *should* so treat it. To show the former he would have to establish some sort of invidious distinction between language-games played (mainly) by philosophers and language-games played by non-philosophers—one which enables us to discount the former as *ipso facto* "mistaken" or "confused." To show the latter, he would have to show that the philosophers' habit of treating "my toothache" as the name of a particular has had disastrous philosophical consequences, and that the only (or the best) way of avoiding these consequences was to break the habit.

I think that the latter project is closest to Cook's actual intention, for he says that when "no two people can feel the same toothache" comes to be called a necessary truth," then "one easily concludes that we cannot know anything about another person's toothaches."²⁵ But this is not an easy conclusion. It does not follow at all unless one grants some further premisses—e.g., (1), or (5) or (8)—all of which, I have argued above, are dubious. A philosopher who refuses to grant any of these further premisses, while holding out for the view that "sensations are private" and explicating "private" in the sense of (P₂), (P₃), or (P₄), is not thereby going to get involved in any of the standard puzzles about other minds, knowledge of the external world, and the like. If he interprets "private" in the sense of (P₁), and thus assents to (8), he will indeed be in trouble. But, once again, we need to notice the distinction between "privileged access" ("privacy" in the combined senses of (P₂) and (P₃)), "unsharability" ("privacy" in the sense of (P₄)), and "incommunicability" (privacy in the sense of (P₁)) and to ask whether, if these distinctions are borne in mind, we cannot get the benefits of Wittgenstein's treatment of private particulars without accepting Cook's paradoxes.

I have now done all I can to show why Cook is

wrong in saying that "sensations are private" is senseless, and thus all I can to argue for the claim labeled (B) in the first paragraph of this article. I turn now to the one labeled (C), and ask whether Cook has produced good arguments for rejecting "I know that I am in pain because I feel it" as senseless. If he can do this, he will have shown that to interpret "privacy" in the sense of (P₂) is wrong-headed.

Cook sets about making this latter claim in an odd way. He says that Wittgenstein wants to get around "Argument A" which goes as follows:

- (i) No one can feel (experience, be acquainted with) another person's sensations.
- (ii) The proper and necessary means of coming to know what sensation another person is having is to feel that person's sensation.
- (iii) Anyone who has a sensation knows that he has it because he feels it, and whatever can be known to exist by being felt cannot be known (in the same sense of "known") to exist in any other way.

Conclusion: No one can know what sensations another person is having.²⁶

One would think that (ii), and the second half of (iii), were so implausible that an attack on them would suffice to get around Argument A. But Cook employs an indirect method, arguing that (ii) and (iii) presuppose that

- (iv) There is a genuine use of the verb "to know" as an expression of certainty with first-person present-tense sensation statements.²⁷

and that this presupposition is false.

The odd thing about this argument is that neither "certainty" nor "certain" appears in (ii) or in (iii), and thus it is hard to see how either could presuppose (iv). Suppose we forget about "certainty" for a moment and ask simply: can we find an account of the meaning of "I know that . . ." according to which "I know that I am in pain" would be sensible? Let us try a conventional analysis, according to which "I know that *p*" is true if *p*, I believe that *p*, and I have adequate evidence for my belief that *p*. Do I have good reasons for believing that I am in pain, when I believe that I am? Two answers suggest themselves:

(A) My reason is that I feel it

and

(B) My reason is that I know the language, and

²⁵ *Ibid.*, p. 309 (318).

²⁶ Cf., *ibid.*, pp. 283–284 (289–290).

²⁷ *Ibid.*, p. 285 (240).

consequently know which states are called "pairs," which "anxieties," which "thrills," etc., etc.

Since Cook has a rather detailed discussion of (A), to which we shall shortly turn, let us concentrate first on (B). At first sight, (B) seems merely to say "I know that I am in pain because I know what pain is," and this does not seem satisfactory. But now let us read (B) as an abbreviation for

(B') My reason is that I am inclined to say "I am in pain," and I know that this inclination, had by one who knows the language, is itself evidence, and indeed conclusive evidence, for the truth of what is said.

Here the person who is asked to justify his belief that he is in pain is calling attention to a convention of the language—the convention that the utterance of, or the disposition to utter, certain first-person present-tense reports are taken as the best possible evidence (and, indeed, evidence which cannot be overridden) for their own truth. In other words, he is distinguishing between the fact that he is in pain and his own belief that he is in pain, and saying that the latter is evidence of the former. One could rule this answer out if one held that

Beliefs can never be evidence for the truth of the proposition believed

but this is false, since other people certainly take my beliefs about my mental states as evidence for their own beliefs about my mental states. So why shouldn't I? Can we justify holding

I can never take the fact that I believe *p* as adequate evidence for the truth of *p*?

Perhaps we can, but I do not see how. It seems to me that we have a choice between letting "privileged" reports count as expressions of knowledge by letting this sort of evidence count, or else ruling them out by ruling out this sort of evidence. I do not know how to make this choice except by looking to the degree of paradoxicalness of the views which will result from each alternative. If we choose the former we have to say "Sometimes the fact that beliefs are held can be adequate evidence for the truth of those beliefs," which sounds a little funny. If we choose the latter we have to say with Wittgenstein "It is wrong to say 'I know what I am thinking'," which also sounds a little funny. Unless we are simply to fall back on our intuitions about

which sounds *more* funny (a result which would, as the recent literature shows, lead to an irresolvable impasse), we have to look further and ask which of the two views leads to the greater *quantity* of funny-sounding views. In other words, we have to look to the further philosophical consequences of taking one alternative rather than the other. Wittgensteinians, by and large, think that a willingness to say "It is wrong to say 'I know what I am thinking'" is the price we have to pay for not having to say the bulk of the funny things which epistemologists in the Cartesian tradition have been wont to say. But, as I remarked above, the familiar Cartesian puzzles would only follow if one accepted, not merely the presupposition of (ii) and (iii) that it makes sense to talk about knowing that one has a certain sensation, but (ii) and (iii) themselves. Once again, we are led to ask why Cook does not attack the premisses rather than their presupposition.

A partial answer to this question will emerge if we go back to the phrase "as an expression of certainty" which Cook drags in, seemingly *ab extra*, in formulating (iv). The reason why he does so is suggested by his saying "Whereas it makes sense to speak of ignorance and knowledge, doubt and certainty, in the case of the stone in the shoe, it does not make sense to speak this way in the case of the man in pain."²⁸ The crucial move here is the implicit suggestion that

(v) Where it does not make sense to say "I doubt that *p*" it does not make sense to say "I am certain that *p*," and conversely.

and that

(vi) Where it does not make sense to say "I am certain that *p*" it does not make sense to say "I know that *p*" (in the sense of "know" intended in (ii) and (iii)), and conversely.

Since one is much more ready to say "I cannot have doubts about whether I am in pain" than to say "It is wrong to say 'I know that I am in pain'" to accept (v) and (vi) is to have persuasive reasons for saying the latter. If Cook and Wittgenstein did not have this move to help them, it is doubtful that they would have been quite so certain that those who believe (iv) are "confused." But this move is at bottom the same as the move which Pitcher attributes to Wittgenstein from

I can identify a certain particular to

²⁸ *Ibid.*, p. 238 (294).

It is in principle possible for me to make a mistake in identifying that particular (a mistake which is not a result of ignorance of the language),

a move which I criticized earlier. Cook is asking us to accept an analysis of "knowledge" according to which so-called "incorrigible knowledge" is not to count as knowledge. In both cases, the argument is clearly "reversible" (in Waismann's phrase). That is, it is not clear whether we should accept the analysis and get rid of the (potentially) embarrassing fact that we possess incorrigible knowledge, or whether we should reject the analysis because it does not cover a certain sort of knowledge (while trying to overcome our embarrassment by other tactics). *Prima facie*, my certainty that I am in pain is an obvious counter-example to (v). If I am told that I cannot render a satisfactory account of the meaning of "certain" that will not entail (v), and of "know" that will not entail (vi), then we are driven back to the question about whether the fact of believing *p* may not sometimes be adequate evidence for *p*. For if it may be, then I am entitled to reject the conjunction of (v) and (vi) as false.

The importance of (v) and (vi) for Cook's view will be evident from the following passage. In discussing the reasons one might adduce in support of "I know it's raining"—e.g., "I'm looking out the window"—Cook says:

What makes it possible to use "I know" here as an expression of certainty is that it would be intelligible for someone to suppose that the speaker is not, in the particular instance, in as good a position as one could want for correctly answering a certain question or making a certain statement. More generally, for "I know that . . ." to be an expression of certainty, it is at least necessary that the sense of the sentence filling the blank allow the speaker to be ignorant in some circumstances of the truth value of statements made by means of the sentence (or equivalents thereof).²⁹

The first sentence of this passage makes clear that Cook tacitly adopts (v). His taking for granted the relevance of "as an expression of certainty" to (ii) and (iii) demonstrates his adoption of (vi). His refusal to consider the possibility that "knowledge of the language" can count as a good reason for knowing that I am in pain can be seen by the fact that his necessary condition for "I know that . . ." being an expression of certainty, as it stands, excludes nothing whatever. For if I do not know what some words in the sentence in question mean,

then *that* is a circumstance in which I can be ignorant of the truth value of statements made by means of the sentence. Since there is no reason to think that everybody knows what "pain" means, this necessary condition, as it stands, is worthless for Cook's purposes. What he obviously intends is that we should take "the speaker" (in both the sentences quoted above) as short for "someone who knows the language." If we make this emendation, then we see that Cook has simply ruled out "knowing the language" as a factor to be taken into account in determining whether someone is in "as good a position as one could want for correctly answering a certain question or making a certain statement." If he did not rule it out, then we could simply reply to him and Wittgenstein: "Certainly I need not suppose that Jones is in as good a position as one could want for determining whether he is in pain; after all, the poor boy is only three, and he still uses 'pain,' 'hurts,' etc., in some very peculiar ways."

Once again, we find ourselves brought back to the alternatives presented earlier: either admit that "knowing the language" counts as part of "being in a good position" (i.e., as a sufficient belief-justifying reason) and accept first-person present-tense pain reports as cases of knowledge, or rule "knowing the language" out and refuse to allow these cases to count as knowledge. The primary reason why Cook prefers the latter alternative is that he thinks that making sensations "objects of knowledge" engenders philosophical perplexities. Specifically, it suggests the question: What is so special about mental states as to make our knowledge of them incorrigible? Now if we answer this by saying "They are directly present to consciousness, and nothing else is," and add on a few other plausible premisses, the resulting epistemological dualism will beget metaphysical dualism (which will beget Idealism, which will beget Neutral Monism, which will beget Logical Positivism) as surely as Sin gave birth to Death. But suppose we answer the question by saying: "There is nothing special about them, apart from a convention that first-person present-tense reports of them are taken as the best possible evidence about them." If we add, with Wittgenstein,³⁰ that when we reach conventions we reach rock-bottom, we will not be tempted to go on to ask "And why do we apply this convention to some things and not others?" We will not fall under the spell of what Pitcher calls the

²⁹ *Ibid.*, pp. 285–286 (291–292).

³⁰ Cf. *The Blue and Brown Books* (Oxford, 1958), p. 24.

"Platonist principle" that differences in degrees of certitude, or of corrigibility, correspond to metaphysical differences in the objects known.

This last suggestion is merely a sketch of a "conventionalist" view of non-inferential knowledge, one which I cannot argue here.³¹ It is inserted here merely to show that there may not be as much reason as Wittgenstein thought to be frightened of the view of sensations are private objects of knowledge. Returning to the business at hand, I shall conclude my discussion of Cook by making one further criticism of detail and then spelling out a more general criticism, which I have sketchily adumbrated above. The criticism of detail concerns his discussion of "the perceptual sense of 'to feel'." Here the defects of Cook's general method of argument seem to me particularly obvious. He wants to argue, in the face of cases like "I feel a pain in my knee," that "Sensation words cannot be the objects of verbs of perception in first-person sentences."³² Clearly, he needs to say that "feel" in the sentence cited, is not "a verb of perception." So he defines "the perceptual sense of 'to feel'" as the sense in which "feel" is used in the sentence "I know it because I feel it." He then says that in such sentences as "I feel a slight pain in my knee when I bend it" the words "I feel" may be replaced by "I have" or "there is" without altering the sense of the sentence. So far so good, and all we need now is an argument to show that when "feel" is used in the perceptual sense such substitution is impossible. But we do not get this. All we get is the fact that in *some* cases of the use of perceptual sense of "feel"—e.g., "I feel a stone in my shoe"—such substitution is impossible. This does nothing whatever to show that satisfaction of the initial definition entails satisfaction of the criterion of the non-substitutability of "There is" for "I feel." One disanalogy does not make a difference of sense, and even if it did, it would not necessarily make the very special difference of sense which Cook needs for his argument. The fact that when we feel stones in our shoes we have methods for checking whether there is one, but that when we feel pains in our knees we do not, hardly shows that it is a "confusion" to answer the question "How do you know that there is a pain in your knee?" by "Because I feel it." Here, as in his comparing "pain" with "build"

rather than with, say, "mountain," Cook uses one analogy or disanalogy between uses of a term to convict his opponents of confusion. If he had confined himself to pointing out that pains, and the way we know about them, differ from stones and the way we know about them, and that something in Argument A, or some other argument with a philosophically undesirable conclusion, depended upon the assumption that we knew about them in just the same way, then all would be well. In certain passages—passages which, to my mind, are the best parts of his article—he does just this. Thus, for example, he notes that "the plausibility of A depends on its seeming to be analogous to something like this: to ascertain whether my neighbor's crocuses are in bloom, as opposed to merely taking his word for it, I must see his crocuses."³³ Later he points up the relevant disanalogy as follows:

It [Argument A] makes out the difference between first- and third-person statements to rest on a matter of circumstance (like being unable to see my neighbor's crocuses) whereas Wittgenstein has made us realize that the difference resides in the language-game itself. The difference does not rest on some circumstance, and therefore Argument A, which purports to name such a circumstance with the words "being unable to feel another's sensations" is inherently confused.³⁴

I think that in the first sentence of this passage Cook accurately characterizes an important part of Wittgenstein's contribution to our understanding of the notion of "privacy." In the second sentence, however, he confuses the fact that we may be inclined to accept the false propositions (ii) and (iii) on the basis of a false analogy between knowing-about-pains and knowing-about-crocuses, with the claim that (ii) and (iii) are not simply false, but are inherently "senseless" or "confused." There is a difference between people being willing to accept a false proposition (or senseless quasi-proposition) *p*, because they accept another proposition (or senseless quasi-proposition) *q*, and *p* entailing or presupposing *q*. Sometimes the phenomena are co-extensive, but it has been the burden of my argument that, in the case at hand, they are not. The claim that the difference between first- and third-person statements about sensations is built into the language-game *can* be accepted by someone who accepts all the premisses of Argument A,

³¹ It is a view which has been elaborated at length by Sellars, and which I try to summarize in "Intuition," *The Encyclopedia of Philosophy* (New York, 1966), vol. 4, pp. 204–212.

³² Cook, p. 289 (295–296).

³³ *Ibid.*, p. 290 (296).

³⁴ *Ibid.*, p. 291 (297).

without his thereby involving himself in logical contradiction, or in making senseless utterances. For, if the foregoing criticisms are sound, Cook cannot get from this difference to the "senselessness of any relevant statements. Nor need he try, since the implausibility of (ii) and (iii) is such as to let us dismiss the argument at once. Granted that it took Wittgenstein to break the spell which these premisses exerted, we need not continue to recite the powerful incantations (such as "It is incorrect to say 'I know what I am thinking'") which Wittgenstein used for this purpose. (It took Hume's paradoxical views to wake 18th-century philosophy from its dogmatic slumbers, but we are fortunate that no misguided sense of gratitude led Kant to reiterate Hume's slogans).

IV. CONCLUSIONS

My discussion of Pitcher and Cook has suggested the following view of Wittgenstein. Wittgenstein wanted to cut Cartesian scepticism off at the roots, and hoped to do so by arguing that the Cartesian picture both of non-inferential and of incorrigible knowledge was incoherent. In particular, he noted that the Cartesian tradition assimilated the language-game played with sensations terms like "pain" to that played with physical-object terms like "tree," and thus produced various false analogies. These analogies inclined philosophers to make various false statements (or to invent a new, specifically philosophical language-game, and then, by combining premisses drawn from this artificial language-games with premisses drawn from more "natural" language-games, to produce arguments which committed fallacies of ambiguity). Pitcher emphasizes the differences between the way in which we *name* trees and pains, and Cook the differences between the way in which we *know about* trees and pains, but both agree that

If sensations are private objects, they can't have names (in the same sense of "name" in which trees have names)

and

If sensations are private objects, we can't know about them (in the same sense of "know" in which we know about trees).

Both, in other words, focus on what Strawson calls Wittgenstein's "hostility to privacy," rather than on what Strawson calls his "hostility to immediacy."³⁵

In the discussion so far, I have admitted that it is possible to construct analyses of "know," and of "name" (or "denote," "refer," or "identify"), which will entail the two theses just mentioned, or alternatively to construct distinctions between different "senses" of these terms which will produce the desired result. But I have argued that Cook and Pitcher present no good reasons for saying that we should adopt such analyses or make such distinctions. I have argued further that both men pick up the problem by the wrong end, that what needs analysis, or distinctions between senses, is "privacy," and that Ayer's distinctions suffice to do this job. I have suggested that it was Wittgenstein's failure to make the distinctions which Ayer makes, and to recognize that an object can be "private" in one of these senses without being "private" in all the others, which accounts for his "hostility to privacy."

In another article, forthcoming, I hope to show that the proper thrust of those portions of the *Investigations* which Malcolm has grouped together as "arguments against the possibility of a private language" is against the notion that we can have knowledge of something distinct from the knowledge that certain propositions are true of it, and against the notion that we can have knowledge of the truth of propositions which are not formulated in language. In other words, I want to argue that what is novel and exciting in these portions of the *Investigations* is the attack on the Cartesian notion of pre-linguistic awareness—the notion that there is a species of awareness which antedates and underlies our coming to be able to justify the utterance of sentences. But this is a separate topic, which I have only had space to hint at in the present article.

Princeton University

Received September 3, 1969

³⁵ Cf. P. F. Strawson, review of *Philosophical Investigations*, *Mind*, vol. 63 (1954), esp. pp. 90 ff.

III. NEGATIVE STATEMENTS

RICHARD M. GALE

AMONG the perennial disputes in philosophy is that between the friends and foes of negative facts or events. The latter argue that negative facts are either reducible to or dependent upon positive facts, while the former, believing in the ontological equality of these two kinds of facts, are ready to rebut these arguments. It is assumed by the disputants that we are able to distinguish between positive and negative facts. This distinction requires that we be able to discriminate between positive and negative statements, since a positive (negative) fact is what is stated by a true positive (negative) statement. If there is no reasonably clear-cut criterion for distinguishing between these two kinds of statements, the dispute between friends and foes of negative facts is like a game played with unmarked dice: in neither case can one determine the winner. The aim of this paper is to examine how such a criterion is to be constructed.

While there is a rich literature on whether negative facts are reducible to or less "real" than positive ones, almost nothing has been done toward devising a criterion for distinguishing between negative and positive statements. This oversight might be explained by the confidence we place in our preanalytic intuitions as to which statements are negative. This confidence may be overconfidence, for, as Frege pointed out, there are many cases in which our intuitions are indecisive, e.g., "This shirt is striped," "Jones is dead," etc.¹ To overcome this indecision we must replace our preanalytic intuitions with a precise criterion for distinguishing between negative and positive statements. Of course, any adequate criterion must yield results which agree with our preanalytic intuitions in those cases in which they are strong and decisive. Without such strong intuitions there would be data neither for launching our task of devising a criterion nor checking its adequacy once it is formulated. Keeping one eye on these intuitions, we shall get on with the task of devising a criterion by first criticizing criteria which have been proposed.

An adequate criterion for distinguishing between positive and negative statements must enable us to determine when a statement is positive and when negative. A criterion for determining when one statement is the negation of another will not enable us to do this. Thus we can bracket as irrelevant any proposed analysis or definition of either the connective " \sim " in terms of its role in inference or the contradictory of a statement. It may be that for the purposes of formulating logical principles and laws there is no need to specify when a statement is negative, as opposed to being negated, but this does not show that there is something illegitimate in our quest for a criterion for a negative (positive) statement. Logic is not the whole of philosophy.

The criterion which most readily comes to mind is a grammatical one. A negative statement is one which is made through the use of a sentence which is grammatically negative, i.e., contains some negative word such as "not" or some negative prefix such as "un." Any such criterion, as Ayer has shown, leads to an absurd consequence, that two different sentences can be used to make statements either that are the same yet differ in quality or that are equivalent but yet different.² Consider the sentences "Hayes is the fastest man" and "There is no other man who is as fast as Hayes." By any grammatical criterion the former is positive and the latter negative. Yet it seems plain that the statements made by the use of these sentences are the same. But then a given statement is both positive and negative, which is absurd. And it would be equally absurd to claim that these sentences are used to make equivalent but non-identical statements, thus allowing one of the statements to be positive and the other negative in relation to these two different sentences. Since grammatically dissimilar sentences can be used to make the same statement it is clear that our quarry is not to be found hiding in a grammar book. This is analogous to the impossibility of giving a grammatical criterion for a statement which describes

¹ "Negation" in *Translations from the Philosophical Writings of Gottlob Frege*, ed. by Max Black and P. T. Geach (Oxford, 1952), pp. 125-126.

² A. J. Ayer, "Negation," *The Journal of Philosophy*, vol. 49 (1952), pp. 797-815, reprinted in his *Philosophical Essays* (London, 1954). All references to this article are from the pagination in the latter book. See pp. 36-37.

an event as being past (present, future). In English these temporal determinations are made through the use of tensed verbs and copula, but in some other languages these same determinations are made through very different grammatical means.³

If grammar cannot supply a criterion for distinguishing between positive and negative statements, maybe psychology can. Many philosophers have claimed that negation signifies a person's mental act of denying, rejecting, or rebutting a statement that is actually made or envisioned as being made by someone.⁴ Accordingly, a negative statement is one which expresses a subject's act of denying some statement, a positive statement not being made in rebuttal of some statement. By making negative facts and events dependent upon a subject's mental act of denial this theory is attractive to those philosophers who cannot get themselves to believe that such facts and events are "objective." Without subjects—persons—there are no acts of denial, and without acts of denial there are no negative facts. If negation really is denial then the opposite of a negative statement is an affirmative rather than a positive statement, the act of denial being opposed to the act of affirmation. That I choose to contrast negative with positive statements indicates that I do not accept the view that negation is denial.

This psychological criterion for distinguishing between negative and positive statements in terms of the mental acts of denial and affirmation just won't do. First, a statement which is negated need not be one which is ever made by anyone or even envisioned as being made. The minor premiss in a *modus tollens* argument negates the consequent of the hypothetical statement in the major premiss, but this consequent statement has not been asserted nor need it be imagined as being asserted. Secondly, a statement that is denied may be either positive or negative, and the same goes for the statement denying it. I might say that it is not the case that it is not raining, herein it being the denied statement that is negative and the denial which is positive,

double negation just being a polite or diplomatic way of making a positive statement.

It would seem that the psychological criterion, at best, enables us to determine when one statement is the denial of another, but it does not indicate which one is negative. This failure also marks Dewey's revision of the psychological criterion, according to which a negative statement is one that rejects a certain fact as not being relevant to the furtherance of an inquiry into some problem, as not being a fact of the case.

Negative propositions . . . represent subject-matters to be eliminated because of their irrelevancy to the evidential function of material in the solution of a given problem.⁵

The question which this theory of negation leaves unanswered is how we determine when some rejected subject-matter or statement is negative rather than positive. I might be engrossed in solving some mathematical problem, and thus the fact that my lamp is blue would rightly be rejected as not being a relevant fact of the case. Still it is a positive fact, regardless of how useless it is in furthering this inquiry.

There are further difficulties with the view that negation is denial which arises from there being conceptual truths pertaining to the negation of a statement which do not hold for the denial of a statement. It is necessarily true that for every statement *s*, either it is not the case that *s* or it is not the case that it is not the case that *s*. If negation is denial then "it is not the case" means the same as "I deny." If so, it should be necessarily true, but is not, that for every statement *s*, either I deny that *s* or I deny that not-*s* (or I deny that I deny that *s*). Moreover, that *s* is false is entailed by that it is not the case that *s* but not by that I deny that *s*. It would seem that negation is part of the factual or informative content of a statement or Thought (in Frege's sense); its function is not a pragmatic one of expressing the propositional attitude of the speaker.⁶

³ For a fuller account see ch. 3 of my book, *The Language of Time* (London, 1968).

⁴ For a defense of this see: Immanuel Kant, *Critique of Pure Reason* (London, 1963), A709, B737; C. Sigwart, *Logic*, vol. I, sec. ed. (London, 1895), p. 119; F. H. Bradley, *The Principles of Logic*, vol. I, sec. ed. (London, 1922), pp. 114-115; Bernard Bosanquet, *Knowledge and Reality* (London, 1892), p. 216; and *Logic*, vol. I, sec. ed. (London, 1911), p. 278; Henri Louis Bergson, *Creative Evolution* (New York, 1944), p. 312; Sidney Hook, *The Quest for Being* (New York, 1961), p. 147; and Hans Regnell, "On Negation and Negative Facts," *Theoria*, vol. 17 (1951), pp. 210-221.

⁵ *Logic: The Theory of Inquiry* (New York, 1938), p. 183.

⁶ Another significant difference between negation and denial is that "I deny" is a performative verb while "It is not the case" is not. To say in appropriate circumstances, "I deny that *s*," is thereby to deny that *s*; however, to say, even in appropriate circumstances (e.g., one in which I have evidence for not-*s*), "It is not the case that *s*," is not for it thereby not to be the case that *s*.

In behalf of the exorcism of negative events and states it is often said that they are not perceived but inferred or judged on the basis of what is actually presented to our senses; e.g., we see that the table is blue, but must infer that it is not green on the basis of what we do actually see.⁷ This can be made to serve as a criterion for distinguishing between positive and negative statements which predicate a sensible property of a spatiotemporal individual. A statement of this kind is positive if the event or state it reports is perceivable, otherwise negative. The major difficulty with this perceptual criterion is that certain events and states which should be classified as negative according to our preanalytic intuitions can correctly be said to be perceived, either with or without a "that" preceding the grammatical accusative of the perceptual verb; e.g., "I see (that) the chair is not blue."⁸ It is no help to say that the negative state of affairs of the chair's not being green, unlike the positive one of its being blue, is not actually imprinted on our retina or on a photographic plate, since this just pushes the problem one stage further back, it being we who must *look* at the retina or photographic plate and determine what it is that we *see*. An appeal to what can be pictured would be equally of no avail.

At this point, having begun to despair at our repeated failures to find a criterion, we might attempt a simple-minded solution. A singular statement is positive if it entails that its subject has the property signified by its predicate, otherwise it is negative. But this gets us nowhere. A singular referential statement could entail that its subject has the property signified by its predicate and still not be positive, since the property in question might be a negative one, such as non-blue. This criterion can work only if we first have a criterion for distinguishing between positive and negative properties or predicates. We feel that blue is a genuine or real property, while non-blue is a mere pretender, getting whatever determinateness it has from its positive complement, blue. But the problem is how to go beyond mere intuition and formulate an explicit criterion for a "real" property, i.e., a positive property.

A. Quinton has attempted to distinguish

between natural and non-natural or arbitrary classes, it being a positive or genuine property alone that determines the former; and he has used this distinction as a basis for determining which of two complementary properties is the positive one.⁹ For Quinton a natural class is determined by a property which is ostensibly teachable, since the things it applies to share some common perceivable property.¹⁰ As a result, a natural class will be open-ended and have members which are representative of the whole.¹¹ Blue determines a natural class, because it can be ostensibly taught. After having a few instances of blue pointed out, and perceiving the property common to these instances, we are able to go on successfully identifying blue things. The class of non-blue things is non-natural, since non-blue cannot be learned in this ostensive manner.

Several things are doubtful about Quinton's criterion. First, we feel that an adequate criterion for distinguishing between positive and negative properties should be logical, having to do with a basic difference between the logic of predicates for these two kinds of properties. Instead we are given a criterion based on the alleged ostensive teachability of positive properties. However, the ostensive teachability of a property depends on all sorts of extra-logical features of a psychological and sociological sort. Ayer asks us to imagine a tribe whose language contains only two color words, "blue" and "eulb," which means a color that is not blue.¹² They attach a mana to blue things and are forbidden to use "blue" except at certain sacred rites. It is reasonable to suppose that the youth of this tribe are first ostensibly taught eulb and then have "blue" defined for them as the complement of eulb. Here it is the negative complement that is ostensibly taught and the positive one which is defined in terms of it.

An even more telling objection is that this criterion really is nothing but a variation on the discredited perceptual criterion, and is subject to all of its difficulties. A positive property alone is ostensibly teachable. Why? Because the things it applies to share some common *perceptible* property, unlike the members forming the extension of a

⁷ For a discussion of this question see G. Buchdahl, "The Problem of Negation," *Philosophy and Phenomenological Research*, vol. 22 (1961), pp. 163-178.

⁸ This point is made by Richard Taylor in "Negative Things," *The Journal of Philosophy*, vol. 49 (1952), p. 33, and Morris Lazerowitz, "Negative Terms," reprinted in *The Structure of Metaphysics* (New York, 1955), pp. 197-198.

⁹ "Properties and Classes," *Proceedings of the Aristotelian Society*, vol. 63 (1957-1958), especially p. 48 ff.

¹⁰ *Ibid.*, pp. 47 and 58.

¹¹ *Ibid.*, pp. 36 and 37.

¹² Ayer, *op. cit.*, p. 49 ff.

predicate signifying a negative property. But, as seen above, it is correct to say that we perceive a negative property. This should not surprise us, since what is perceived is so dependent upon linguistic, psychological, and sociological facts involving the perceiver, as well as features of the immediate context of perception.

Ayer has claimed that in respect to a pair of complementary properties there is no criterion by which it can be shown that one of them alone is a "real" or "genuine" property. His denial is based on acceptance of the principle of significant contrast as the criterion for a meaningful predicate. If one member of a pair of complementary predicates satisfies this principle so does the other.

The fact is that every significant predicate has a limited range of application. Its correct use is determined both by the fact that there is a set of occasions to which it applies and by the fact that there is a set of occasions to which it does not apply. This being so, it must always be possible to find, or introduce, a predicate which is complementary to the predicate in question, either in the wide sense, as applying to all and only those occasions to which it does not apply, or, in a narrow sense, as applying to all and only the occasions of this sort that fall within a certain general range.¹³

"Blue" applies to blue things and to nothing else, while "eulb," interpreted as the narrow complement of "blue," applies to things which are a color that is not blue and to nothing else. What is common to all things to which the former applies is that they are blue and in respect to the latter that they are a color that is not blue. Neither signifies a more genuine or real property or universal than the other.

Ayer makes quick work of Cook Wilson's attempt to show that a negative term does not signify a "true" universal. According to Wilson:

If *Aness*, *Bness* and *Cness* are all different in the sense that every *A* is not *B*, every *B* not *A*, &c., it would follow as before that *Cness* was a species of not-*Aness* and at the same time a species of not-*Bness*. But, if a universal is a differentiation of two different universals, either these two are related themselves as true genus and species or they mutually involve one another so that each necessitates the other. Thus though not-*Aness*

and not-*Bness* exclude one another, either one is a species of the other, or each involves the other, but both these alternatives are self-contradictory.¹⁴

Let *A*, *B*, and *C* be respectively blueness, greenness, and redness. These universals are species of color and are different from each other in the required sense, since every blue thing is not green, etc. The complements of blueness and greenness are respectively non-blueness and non-greenness. Redness is a species of each of these, since that something is red entails that it is non-blue and that it is non-green. According to Wilson's criterion for a true universal, since redness differentiates both non-blueness and non-greenness there must be some relation of necessitation between them; that something is non-blue entails and/or is entailed by that it is non-green. Plainly these relationships fail to hold, since if an object is green (blue) it is non-blue (non-green), but cannot without contradiction be non-green (non-blue). In regard to true universals—positive ones—this relation of necessitation holds; e.g., navy blueness differentiates blueness and coloredness, but blueness necessitates coloredness.

To this Ayer replies:

Now, if we were obliged to hold that either of the complementary predicates to "blue" and "green" entailed the other, let alone that each entailed the other, we should certainly be involved in contradiction; for it can easily be shown that if this were so the same thing might be, in the one case, both blue and not blue and, in the other, both green and not green. But why should we have to make any such assumption? No reason has been given for supposing that just because the complementary predicates to "blue" and "green" have a number of other predicates, such as "brown" and "red" and "yellow," subsumed under them both, they must therefore stand in any relation of logical necessitation to one another. And if it be alleged that this is true of all universals that have any common specification, the answer may be that it is just such cases as these that provide a counter-example.¹⁵

This way of dismissing Cook Wilson's criterion completely misses the point. In the paper from which the preceding quotation is taken, Ayer is trying to formulate a criterion for a negative statement. He very likely is right in challenging Wilson's

¹³ *Ibid.*, p. 48. I assume that Ayer's principle of significant contrast requires only that for a predicate to be meaningful it must be logically possible for both it and its complement to apply to something, not that both actually do have an application. The problem with assuming the latter is explored in my article, "Hook's Views on Metaphysics" in *Sidney Hook and the Contemporary World*, ed. by P. Kurtz (New York, 1968).

¹⁴ John Cook Wilson, *Statement and Inference* (Oxford, 1926), vol. I, p. 254.

¹⁵ Ayer, *op. cit.*, pp. 52-53.

criterion as a criterion for a "true" universal. Wilson may be guilty of linguistic imperialism here, of arbitrarily stipulating a definition of "universal," not a very serious offense since this is a term of philosophical jargon. This is irrelevant, however, to whether this is an adequate criterion of a negative universal or property, which is what Ayer is supposed to be seeking. What shall be shown is that Ayer's own criterion, which is based on the level of specificity between complementary predicates, fails and that something like Wilson's criterion can work.

It must be borne in mind that Ayer is seeking a criterion for a negative *statement*. This comes out in his rejection of any grammatical criterion based on determination of when a *sentence* is negative.

When philosophers contrast affirmative with negative statements, the distinction with which they are concerned applies not to the grammatical form of different sentences but to the different ways in which they are used.¹⁶

After showing that a psychological, as well as an ostensive teachability, criterion will not work, he goes on to reject any attempt, such as that of Cook Wilson above, to distinguish between negative and positive predicates by showing that only the latter signify a real universal or property. In respect to a pair of complementary predicates, if either satisfies the principle of significant contrast—Ayer's criterion for a meaningful predicate—so does the other.

Although both of a pair of complementary predicates can be meaningful, and thus express a real property, Ayer claims that in certain cases one of the pair is negative, and, what is more, can be shown to be such by a criterion that yields results generally in agreement with our intuitions. On the basis of this criterion it can be determined when a statement is positive and when negative. The following is Ayer's specificity criterion:

Let us say of any two singular referential statements *S* and *S'*, that is, statements which refer to a particular individual and ascribe some property to it, that *S* is a specifier of *S'* if and only if *S'* is not a component of *S*, *S* entails *S'*, and *S'* does not entail *S*. And let us say that a singular referential statement *S* is absolutely specific, with respect to a given language *L*, if there is no statement expressible in *L* which is a specifier of *S*. Further let us say that a statement has the first degree

of specificity, with respect to a language *L*, if it is not absolutely specific but has no specifiers expressible in *L* which are not themselves absolutely specific, that it has the second degree of specificity if it is not absolutely specific, or specific to the first degree, but has no specifiers which are not themselves absolutely specific or specific to the first degree, and so on. Then among complementary pairs of singular referential statements it may happen that one member of the pair has a higher, that is, a less, degree of specificity than the other. In that case the more specific statement may be said to be affirmative and the less specific to be negative.¹⁷

Notice that Ayer is not saying that a statement is affirmative (positive in my terminology) in relation to some statement of which it is a specifier and negative in relation to some other statement which specifies it. This would yield the strange consequences that the statement that *a* is blue is positive in relation to the statement that *a* is colored but negative in relation to the statement that *a* is azure. Rather, according to this criterion, it is the case that, in English, that *a* is blue is positive, since it is more specific than is the statement predicating of *a* the complement of blue—non-blue.

Once we have determined the class of positive singular referential statements by this criterion the quality of quantified statements can be determined as follows:

An existential statement may be said to be affirmative if and only if it is entailed by an affirmative singular referential statement. A universal, or particular, statement may be said to be affirmative if and only if the singular referential statements which are obtained from it by giving values to its variables are affirmative. The class of negative statements, with respect to a given language, may then be held to consist of all and only those statements that have affirmative statements for their complementaries.¹⁸

By the use of the principle of quantifier negation certain quantified statements can be made through the use of either a grammatically affirmative or negative sentence. This will not result in ambiguous classifications, since there is no way of instantiating a negatively quantified statement nor any way of deriving such a statement by means of generalization from a singular statement.

Does Ayer's criterion work? In answer to this it will be argued that: (1) his criterion should not be

¹⁶ *Ibid.*, p. 38. We shall see that Ayer's criterion, because it is relativized to a language, betrays his own declared objective.

¹⁷ *Ibid.*, p. 52. Ayer uses "affirmative" instead of "positive." My reasons for preferring the latter term have been given in my argument against negation being the same as denial.

¹⁸ *Ibid.*, p. 63.

relativized to a language; (2) it is not possible to compare the level or degree of specificity of singular referential statements predicating complementary properties of the same individual without reference to a hierarchical system of classification containing these properties; (3) his vertical criterion based on the level of specificity, even when applied to complementary properties within a hierarchical system, yields counter-intuitive results, which can be avoided by a horizontal criterion concerned with connections between properties on the same level of specificity within a system. Various limitations in this horizontal criterion will then be indicated; and, finally, it will be shown that there are versions of Cook Wilson's criterion, interpreted as a criterion for a positive rather than a "true" universal or property, which succeed in distinguishing between negative and positive properties and are not subject to these limitations.

(1) According to Ayer's criterion a statement is positive (negative) only in relation to some specific language. He writes:

Thus, while for the members of our imaginary tribe the statement that an object was blue would not be more specific than the statement that it was eulb, to say in English that an object is blue is to make a more specific statement than to say that it is not blue; for whereas the statement that an object is not blue is specified by such statements as that it is red, or that it is green, or that it is colourless, the statement that it is blue is not specified by any statement at this level. Accordingly the statement that the object is blue counts, with respect to the English language, as affirmative by this criterion, and the statement that it is not blue as negative.¹⁹

By so relativizing his criterion to a specific language, Ayer is inconsistent with his own objection to a grammatical criterion. According to this objection, a grammatical criterion can yield ambiguous results, a given statement being both positive and negative, depending upon whether it is made through a positive sentence or a synonymous negative sentence. Ayer assumes that a statement cannot have more than one quality; if it is negative, it is not positive (or neither the one nor the other). But, Ayer tells us, a statement could be negative if made by the use of a sentence in one

language, e.g., "*a* is not blue" in English, but neither positive nor negative if made by the use of a sentence in a different language, e.g., "*a* is eulb" in the language of his tribe—Tribese. If the seeming contradiction of a statement's having incompatible qualities can be explained away by relativizing a statement's quality to a language, then relativizing its quality to a sentence should be an equally good, or bad, way out.

More important than the charge of inconsistency, which is only *ad hominem*, is the fact that Ayer's attempt to relativize a statement's quality to a language involves a distorted conception of a statement. Judging from what Ayer says on p. 38 of his article, a statement has to do with the way in which a sentence is used. This seems to fit the definition of a statement as the what-is-said (-believed, -judged, etc.), which is indicated through an *oratio obliqua* construction. It will be left an open issue whether such intentional accusatives can be reduced to acts of stating (etc.) or properties of such acts. It will be assumed that if two sentences are used to make the same statement the statements they make have the same entailment relations, i.e., they entail and are entailed by the same statements. Granting this as a necessary condition for statement identity, it can be shown, *pace* Ayer, that one and the same statement cannot differ in its quality relative to different languages. For this to happen a given statement must differ in its level or degree of specificity relative to its complementary statement in different languages. To ensure that this cannot happen all we need show is that a statement cannot differ in its level of specificity in different languages. It will be seen that our assumption concerning statement identity ensures that this cannot happen.

Ayer's example of a statement which differs in its specificity level in different languages is the statement which is made both through the use of the English sentence "*a* is not blue" and the Tribese sentence "*a* is eulb." It is claimed that these sentences are used to make statements which differ in their level of specificity, since the speakers of Tribese, because of its paucity of color words, cannot, as can speakers of English, express statements, e.g., that *a* is red (green, etc.), which are

¹⁹ *Ibid.*, p. 62. By placing colorless on the same "level" as red, green, and blue, Ayer is taking it to mean something like transparent or white, both of which are specifiers of color. Colorless as the complement of colored means not subject to color determinations. But if Ayer is using colorless in this way he is inconsistent when he later says (on p. 63) that one of the paradoxical consequences of his specificity criterion is that "the statement that an object is coloured will have to count as negative: for it is less specific than the statement that the object is colorless." For his criterion applies only to complementary predicates. If colorless is a specifier of color, it cannot also be its complement.

specifiers of that *a* is not blue. Forgetting, but only for the moment, about infimae species of blue, red, etc., it follows that the statement that *a* is not blue is absolutely specific in Tribese but of the first degree of specificity in English.

Ayer faces a dilemma here. Either the speakers of Tribese and English make the same statement or they do not. If they do not, it is not one and the same statement which has different levels of specificity in different languages; and, if they do, then, according to our assumption for statement identity, these statements have the same entailment relations and therefore the same degree or level of specificity.²⁰ On either horn it turns out that Ayer fails to produce an example of a *single* statement having different levels of specificity in different languages. And if such an example cannot be produced, a statement cannot differ in quality relative to different languages. Of course, we cannot identify or refer to a statement except by using a sentence in some language, but in using this sentence we are not referring to it. In saying, "the statement that *s*," I refer to the statement that *s*, not to the sentence '*s*'.

(2) Even when Ayer's specificity criterion is not relativized to a language there are problems. One concerns how we are to compare the respective levels of specificity of *wide* complements. Consider in this connection the pair of wide complementary properties, number, and non-number. That *a* is a number is of at least the sixth degree or level of specificity, since it has specifiers belonging to six different levels of specificity; in inverse order of specificity some of them are that *a* is a real number, a rational number, an integer, a positive integer, a positive odd integer, the number seven. Non-number, being the wide complement of number, applies to everything that is not a number, e.g., an organism. Organism is a property of at least the third degree of specificity, since it is specified by statements on three different levels; some of them are that *a* is an animal, a man, a Caucasian. This makes it appear as if non-number is of the fourth degree of specificity, and therefore positive, since it is more specific than its complement, number. If this counter-intuitive result should bother us we no doubt could find some specifier of non-number, such as *thought of a number*, which would make non-

number appear to be of a higher degree of specificity than number, and therefore the negative member of the pair. Obviously, something has gone wrong, for a given property cannot be both positive and negative.

The way out of this difficulty is to restrict Ayer's specificity or vertical criterion to *narrow* complements. This narrowing of the universe of discourse requires that the complementary properties in question be species of some common genus. Blue and color that is not blue are narrow complements, since they both specify color. We can now state Ayer's vertical criterion as follows:

- (V) A property is positive if and only if it is more specific than its narrow complement, i.e., is specified by properties on fewer horizontal levels.

A property is negative if less specific than its narrow complement, and neither positive nor negative if on the same level of specificity as its narrow complement. This criterion presupposes a classificatory system containing the narrow complements to be tested. Such a system is made up of both a vertical and horizontal dimension. The vertical dimension is determined by relations of being a specifier of, which hold between the infimae species and the species on the next higher level, and so on up to the summum genus of the system. The degree or level of specificity of a property is determined by its position in this vertical dimension. A horizontal level of the system is determined by a group of incompatible species having the same degree of specificity. It will be seen that this reliance on a classificatory system is a weakness in criterion (V).

(3) There are problems with criterion (V). First, if a genus logically could have only two species then each of these is the narrow complement of the other. But then there is a pair of narrow complements neither of which is positive, since neither is more specific. Dry and damp, odd and even, are examples of such complements. Therefore, to say that *a* is even (odd) or that *a* is damp (dry) is not to make a positive statement (nor a negative one), which runs counter to our intuitions that they are positive statements.²¹ This difficulty can be met by placing an *ad hoc* restriction on (V), so that it applies only to properties that belong to a classifica-

²⁰ Which horn we pick is based on what criteria we adopt for determining when two persons have made the same statement or said the same thing. A tough criterion would require that each person can verbally formulate all of the statements which the other claims is entailed by or entails this statement. According to this criterion the speakers of Tribese and English do not make the same statement.

²¹ Ayer concurs. See *ibid.*, p. 63.

tory system which admits the logical possibility of there being more than two species on any one of its horizontal levels.

Another kind of counter-intuitive result, which cannot be met by any *ad hoc* restriction, is the following. Suppose there is a genus *G* which has *C*, *D*, and *E* for its species, but *C* alone has species, and these species have infimae species. Therefore, that *a* is *C* is specified by statements on more horizontal levels—two—than its narrow complement that *a* is a *G* that is not *C*, it being specified by statements on only one level. An example of this would be where *G* is color and *C*, *D*, and *E* are blue, green, and red, respectively, and blue alone has species, such as azure, which themselves have infimae species, such as kazure (a determinate shade of azure). The statement we make by using the sentence, "*a* is red (green)," is not exactly the same, although a close cousin of, the one which we ordinarily make, since these differ in their entailment relations. That *a* is blue is negative, since its narrow complement is specified by statements on fewer horizontal levels. It is not logically impossible that colors should be organized in this way, it being a contingent fact that there are the species of color there are and that these species have the species they have, and so on. Even if colors were to have the strange stratification envisioned above, we would not want to say that blue is a negative property.

This counter-intuitive result can be avoided by replacing the vertical criterion (V) with the following horizontal one:

- (H) A property is positive if it is not specified by any property on the same horizontal level as its narrow complement, otherwise it is negative.²²

This horizontal criterion meets both of the preceding objections. First, both damp and dry (odd and even) will now check out as positive, since none of these properties is specified by a property on the same horizontal level as its narrow complement. Secondly, regardless of whether there are determinate shades of the different species of color and determinate shades of these determinate shades, and so on, that *a* is a color that is not blue is specified by statements containing predicates on the same

horizontal level as its narrow complement, e.g., that *a* is red.

The horizontal criterion (H), like the vertical criterion (V), presupposes a genus-species or determinable-determinate classificatory system, because it assumes that we already know when two properties are on the same level of specificity within a system. By presupposing a classificatory system, these criteria face certain difficulties. One is due to the fact that not all singular referential statements have predicates signifying properties which fit a genus-species or determinable-determinate type of classification; e.g., "Smith is (is not) a member of the Elks," "The carburetor is (is not) the defective part of the automobile."²³ Being a member of the Elks (being the defective part of the automobile) is not a property contained within a genus-species classification. That it is not seen by the unanswerable nature of the question, "What property on the same level of specificity as being a member of the Elks is incompatible with it?" Our intuitions tell us that statements which refer to some set or whole and say that some member or part of this set or whole does not have a certain property are negative, but neither criterion (V) nor (H) can be utilized to show that they are.

Another difficulty with both (V) and (H), which will be discussed in section (4), is that by presupposing a classificatory system they presuppose the very distinction they are trying to explicate. Just why will become clearer after we have formulated some versions of Cook Wilson's criterion which do not presuppose a classificatory system—a task to which we now turn.

(4) What is distinctive about the Cook Wilson criterion is that it is based solely on a difference in the kind of logical relations into which positive and negative properties enter, without presupposing any classificatory system. It will be shown that: (i) there are several viable versions of a Cook Wilson type criterion; and, (ii) they are superior to both the vertical criterion (V) and the horizontal criterion (H).

(i) According to Cook Wilson, "if a universal is a differentiation of two different universals, either these two are related themselves as true genus and species or they mutually involve one another so that

²² Ayer, without indicating it, shifts from his vertical criterion to a horizontal one when he says on p. 62: "Whereas the statement that an object is not blue is specified by such statements as that it is red, or that it is green, . . . , the statement that it is blue is not specified by any statements *at this level*." My italics.

²³ H. H. Price's contribution to the symposium on "Negation" in the *Proceedings of the Aristotelian Society*. Supplementary Volume IX (1929), pp. 97-111, is very helpful on this point.

each necessitates the other.²⁴ This can serve as the basis for the following criterion for a negative property:

(W) A property P is negative if and only if there is a property P_1 such that:

- (a) P and P_1 are both specified by a property P_2 ; and
- (b) neither P entails P_1 nor P_1 entails P .

A property that fails to satisfy (W) is positive. To say that a property F entails a property F_1 means that that something is F entails that it is F_1 . To say that a property F specifies a property F_1 means the following: (a) F entails F_1 ; (b) F_1 does not entail F . A property may specify another without being either a determinate or species of it. This is due to conditions (a) and (b) being necessary but not sufficient for a determinate-determinate or genus-species relation.²⁵

Unfortunately, (W) does not work, since it is satisfied by properties that obviously are positive. E.g., let P be blue, P_1 be non-azure, and P_2 be navy blue. P_2 specifies both P and P_1 , and P does not entail P_1 nor does P_1 entail P . P_1 is itself negative, and it might be thought that (W) could be protected against this kind of counter-example by restricting P_1 to positive properties; however, this would make (W) viciously circular. This circularity can be avoided by restricting P_1 to a property of the same quality as P . Being of the same quality as, no doubt, is epistemically or psychologically less primitive than being positive (negative); but there is no reason why we cannot take it as being logically primitive and devise a criterion for distinguishing between negative and positive properties in terms of it. In other words, it is being assumed that we have the concept of being of the same quality as, and, as a result, can segregate properties into two groups, each member of one group being of the same quality as every other member of that group and no member of one group being of the same quality as any member of the other group. A criterion will now be formulated for determining which group contains the positive properties and which the negative. It is as follows:

(W)₁ A property P is negative if and only if there is a property P_1 such that:

- (a) P_1 is of the same quality as P ;

- (b) P and P_1 are both specified by a property P_2 ; and
- (c) neither P entails P_1 nor P_1 entails P .

This criterion is an improvement. Blue no longer turns out to be negative, for while it is true that blue and non-azure are both specified by navy blue and there is no entailment relation between blue and non-azure the latter is not of the same quality as blue. There still remains a counter-example to (W)₁. Imagine that there is a property *aoi* which denotes a part of the spectrum that overlaps the part denoted by blue. Let us call the part of the spectrum that is both blue and *aoi* *flue*. *Flue* specifies both blue and *aoi*, even though blue and *aoi* are of the same quality and stand in no entailment relation, thereby satisfying (W)₁. This counter-example can be met by requiring P_2 to be of a different quality from P :

(W)₂ A property P is negative if and only if there is a property P_1 such that:

- (a) P_1 is of the same quality as P ;
- (b) P and P_1 are both specified by a property P_2 which is not of the same quality as P ; and
- (c) neither P entails P_1 nor P_1 entails P .

Any property that fails to satisfy some condition of (W)₂ is positive.

There still remain counter-examples to (W)₂, which can be met by restricting the sort of properties that can be substituted for the property variables, P , P_1 , and P_2 . Here is one example. Let the property to be tested P be blue. There is a property P_1 —square—such that: (a) P_1 is of the same quality as P ; (b) P and P_1 are both specified by a property P_2 —blue-and-square-and-non-red—which is not of the same quality as P ; and (c) neither P entails P_1 nor P_1 entails P . The reason why P_2 differs in quality from the positive property P is that a conjunctive property, at least according to my intuitions, is negative if any one of its conjuncts is; a disjunctive property is positive if any one of its disjuncts is. This is analogous to a conjunction being contingent if any one of its conjuncts is and a disjunction being necessary if any one of its disjuncts is.

Admittedly, blue-and-square-and-non-red is a strange sort of a property, if it is one at all. How can

²⁴ Cook Wilson, *op. cit.*, p. 254.

²⁵ For an indication of what additional conditions are required for a definition of these two relations see John Woods, "On Species and Determinates," *Nous*, vol. 1 (1967), pp. 243-254.

we restrict the properties over which the property variables of $(W)_2$ range so as to eliminate this kind of a counter-example? One way is to restrict the range of properties to "atomic" properties, i.e., a property which contains a constituent property that entails every one of its constituent properties. Red, for example, is atomic; for while it is a conjunction of red and colored, its former constituent entails every one of its constituent properties. This restriction eliminates the above counter-example, since blue-and-square-and-non-red obviously is not atomic. This, however, is not a good way of restricting $(W)_2$, since, while it would work for determinate properties, e.g., red, it would be inapplicable to properties which are a species of some genus, e.g., square is a conjunction of quadrilateral-and-equiangular-and-equilateral, no one of whose constituents entails every one of its constituents. This unwanted result can be avoided if we restrict $(W)_2$ instead to "qualitatively homogeneous properties," i.e., properties all of whose constituents are of the same quality. Again, we have had to make use of our logically primitive notion of being the same quality as. It is apparent that blue-and-square-and-non-red is not qualitatively homogeneous. I do not think it would be accurate to call such a property "qualitatively indefinite," since, as indicated above, a conjunction one of whose conjuncts is negative is itself negative. Restricted in this way, I believe that $(W)_2$ works.

There are simpler alternatives to $(W)_2$ which also are based on a difference in the kind of logical relations sustained by negative and positive properties, without a classificatory system being presupposed. Condition (b) of $(W)_2$ can alone serve as a criterion for a negative property, since a negative property specifies only another negative property whereas positive properties specify both positive and negative properties. By once again making use of the primitive notion of having the same quality as, this can be expressed as follows:

$(W)_3$ A property P is negative if and only if it specifies no property which is not of the same quality.

A property that does not satisfy some condition of $(W)_3$ is positive. There are specification relations going up a genus-species classification, e.g., from

man to animal, but there also are specification relations going down this classification when we insert for each property its wide complement, e.g., from non-animal to non-man. But positive properties, in addition to specifying other positive properties, also specify negative properties, i.e., properties of different quality. E.g., blue specifies, in addition to color, non-red, but non-red, as the wide complement to red, specifies no positive property, i.e., property differing in quality from it.²⁶ Again, it is necessary to restrict $(W)_3$ to qualitatively homogeneous properties; otherwise, non-red color would have to count as positive, since it specifies color, which differs from it in quality.

Another difference between positive and negative properties in respect to their logical relations, which also can serve as a criterion, is the following:

$(W)_4$ A property P is negative if and only if there is no property of the same quality as it with which it is incompatible.

To say that two properties are incompatible means that it is logically impossible for them to be co-instantiated. Red is incompatible with blue, which is of the same quality as it, but there is no property of the same quality as non-red with which it is incompatible. $(W)_4$ also must be restricted to qualitatively homogeneous properties; for non-red color, an obviously negative property, is incompatible with some other property of the same quality—non-colored.

It might be wondered how criteria $(W)_2$ – $(W)_4$ would work for a property such as striped or speckled. To say of a surface that it is speckled means that parts of its surface are of one color and its other parts are not of this color. Thus, this property is qualitatively heterogeneous, and as a result these criteria do not apply to it. What we can do is to apply our criteria to each of its atomic constituents. It turns out to be a conjunctive property, and since one of its constituents is negative it is negative. My intuition concerning the quality of conjunctive (disjunctive) properties one of whose conjuncts (disjuncts) is negative (positive) are not too strong, and there may be exceptions to the rule that a conjunctive (disjunctive) property with a negative (positive) conjunct (disjunct) is negative (positive). We might eschew trying to

²⁶ It might be contended that there are properties not of the same quality as non-red which it specifies, viz., being an entity, being something, being self-identical with itself, etc. These "properties" will not be counted as properties, since they logically must have a universal extension. We shall also not countenance as properties any property that logically must have a null extension.

classify qualitatively heterogeneous properties, and instead confine ourselves to the classification of their atomic constituents.

(ii) Some viable versions of a Cook Wilson type criterion, viz. $(W)_2$ – $(W)_4$, have been presented, and it will now be shown that they are superior to the vertical criterion (V) and the horizontal criterion (H). First let us consider how adequately these criteria handle properties, e.g., damp and dry, which specify a genus which logically could have only two species. It has been shown that (H) is superior to (V) in this regard, since (H) alone can apply to these kinds of properties. Criteria $(W)_2$ – $(W)_4$ are at least the equal of (H), since they too require us to classify these properties as positive. E.g., damp is incompatible with a property of the same quality—dry—thereby failing to satisfy some condition of $(W)_4$; and, moreover, it fails to satisfy some condition of $(W)_3$ since it specifies a property of a different quality, e.g., non-number.

While criteria (H) and $(W)_2$ – $(W)_4$ are right in general to classify such properties as positive, there are “teleological” counter-examples in which one of the two logically possible species of a given genus is considered to be natural, fit, or proper to the kind of things that instantiates it. When a thing of this kind does not have this “natural” property it suffers a “privation” in the sense of Aristotle and Aquinas. In such teleological cases a statement predicating the narrow complement of this natural property is to be classified as negative. E.g., under the genus of organism having eyes there are only two possibilities—having the ability to see and being blind. Yet to say that a man is blind is to make a negative statement according to my intuitions, the reason being that we think it fit or natural for a man to be able to see. Similarly, to say that a man is dead is to make a negative statement according to my intuitions, since life seems to be a more natural and desirable state than death.

The superiority of $(W)_2$ – $(W)_4$ over (V) and (H) rests on the fact that the former alone do not presuppose a genus-species type classificatory system which contains the property to be tested. Two difficulties arise for (V) and (H) over the presupposition. First, as already indicated, their generality is limited, because they cannot be applied to properties which do not fit a genus-species or determinable-determinate classification, e.g., being a member of the Elks. But criteria $(W)_2$ – $(W)_4$ can successfully handle such cases; e.g., being a member of the Elks is positive since it fails to satisfy $(W)_4$, being a member of the Elks being incompatible

with being a number. It also fails to satisfy $(W)_3$, since it specifies being a non-number, which differs from it in quality.

The major difficulty with presupposing a genus-species type classification is that it seems to presuppose the very distinction it is supposed to explicate. A genus-species type classification itself presupposes a distinction between positive and negative properties. The reason is that such a classification cannot be determined by negative properties, only by positive ones. The species of a given genus must be mutually incompatible, otherwise there would be cross-classifications. A negative genus, according to $(W)_3$, can have as a species, be specified by, only another negative property. But, according to $(W)_4$, a negative property is not incompatible with any other negative property, i.e., property of the same quality as itself. Thus, from $(W)_3$ and $(W)_4$ it can be deduced that no two species of a negative genus are incompatible. And for this reason negative properties cannot determine a genus-species classification. Of course a negative genus can have species which are positive properties, since a positive property can specify a property of a different quality. But these positive species need not be incompatible. E.g., non-red has among its positive species justice and action, but they are not incompatible since something could be a just action.

Not only do $(W)_2$ – $(W)_4$ not presuppose a genus-species classification, but they explicate the distinction between positive and negative properties in terms of those features of positive properties that make them alone fit for determining such a classification. A philosophically illuminating criterion for distinguishing between positive and negative properties must not just yield results in agreement with our pre-analytic intuitions; if this were all that is required, a perfectly adequate criterion would be that a property P is positive if and only if it is of the same quality as red. In addition, this criterion must indicate those features of positive and negative properties which explain why the former play a role in the development of systems of knowledge that the latter cannot, why negative properties are epistemically of no significance. One of the jobs we want properties to do is to help us in constructing classificatory systems of the genus-species sort, but this is just what negative properties cannot do. The great value of $(W)_2$ – $(W)_4$ is that they explain why negative properties are unfit for this purpose. One can sympathize with a Cook Wilson who denies that negative properties

are "real" or "genuine" universals. Plato's science of dialectics, I suspect, consists in constructing hierarchical classifications of the Forms, his relation of "mingling" or "blending" being that of being specified by. There could be no such a science if we allow negative Forms into the realm of Being.²⁷

University of Pittsburgh

Received April 22, 1969

²⁷ Much of the work on this paper was done during the spring term of 1968 while I held a Charles E. Merrill Fellowship at the University of Pittsburgh. I profited immensely from discussions with the following: Alan Ross Anderson, Nuel D. Belnap, Jr., Joseph Camp, José Alberto Coffa, Charles Hamblin, Philip Quin, Nicholas Rescher, Michael Rohr, Robert Schultz, Thomas Schwartz, and Martin Tweedale.

IV. IN DEFENSE OF THRASYMACHUS

T. Y. HENDERSON

CRITICS and commentators on the dispute between Socrates and Thrasymachus which occurs in Book I of the *Republic* come closest to unanimity on two points: that Thrasymachus' view is not consistent throughout the dispute and that Socrates' arguments are fatal to Thrasymachus' position. Not all commentators by any means, however, are in agreement as to just what position Thrasymachus is defending, or on what points, precisely, he is inconsistent. Nor is there general agreement as to how Socrates' arguments refute his position.

The dispute between Socrates and Thrasymachus centers around two major points of disagreement: (1) what the nature or essential quality of *justice* is, and (2) whether the just or the unjust life is the *best* (in the sense of most profitable) life for men to live. It is not clear that every commentator agrees that Plato was even aware of the different nature of these two questions, but in the text he has Socrates postpone attention to (2) until he has silenced Thrasymachus on question (1). An examination of the arguments involved in both of these questions would render this discussion much too lengthy, so I shall concentrate my attention almost entirely upon question (1). I propose to offer an interpretation of Thrasymachus' view of the nature of justice and the just life which, in its essential points, is consistent throughout Book I. It is true that Socrates forces him to reformulate his position in a couple of instances, but in so doing, I shall argue, Thrasymachus does *not* change his basic, or original, position.

Not only do I believe that Thrasymachus is consistent in essentials throughout the dispute with Socrates over the nature of justice, I shall also argue that Socrates' most vigorous attacks fail completely to refute, or even seriously to damage, Thrasymachus' position. Yet I want to be very clear that my interpretation of Thrasymachus is not meant to be a "contemporary reformulation," in the sense of claiming that this is the way Thrasymachus *should* have argued; nor am I claiming either that Plato deliberately had Socrates misunderstand Thrasymachus' position, or that Plato himself really didn't

understand what Thrasymachus was saying. My claim is stronger: I believe that the interpretation I shall give is the position Thrasymachus held, that Plato understood it in this way, and that in the dialogue Socrates addresses himself to it directly. If his arguments fail to refute Thrasymachus, as I think they do, it is *not* because the disputants are arguing at cross-purposes, but rather because Socrates' arguments are defective.

* * *

In regard to the question of the essential nature of justice, Thrasymachus is most often accused of inconsistency on the following points:

(1) He offers two formulations of his definition of justice, which I shall refer to hereafter as *Ja* and *Jb*:

Ja: "I declare that justice is nothing else than that which is advantageous to the stronger."¹

Jb: "...justice and the just is really the good of another, the advantage of the stronger who rules, but the self-inflicted injury of the subject who obeys; that injustice is the opposite, and rules those very simple just souls; that the governed serve the advantage of the stronger man, and by their obedience contribute to his happiness, but in no way to their own (343c2-7)."

The alleged inconsistency here is that *Ja* can be interpreted as implying that whenever a strong man acts justly this will result in benefit or profit to himself; whereas *Jb* can be viewed as implying that whoever acts justly, the strong or the weak, such action never results in profit to the agent, but rather to someone else.

(2) Thrasymachus maintains that it is most often the case that the strongest man in the state is the ruler, and that the rulers in every type of state make laws which are to their own advantage (338e1-5). Thus according to *Ja*, if the subjects obey the law they should be acting justly. But Socrates points out that rulers are capable of making mistakes in their legislation and thus might mistakenly enact a law which was not to their own advantage (339c1-38). In such case the subjects would be acting justly if

¹ *Republic*, 338c1-2, tr. A. D. Lindsay (London, 1935). All quotations herein are taken from this translation.

they obeyed the law, but their actions would not result in benefit to the stronger, thus contradicting *Ja*.

(3) Following Socrates' appeal to the analogy of other arts to the "art of ruling," Thrasymachus refuses to draw the conclusion which Socrates believes the argument entails, but instead appeals to another kind of "art," that of shepherding. Socrates himself then accuses Thrasymachus of inconsistency, in that in speaking as he does of shepherds, he fails to follow his own recommendation to discuss only the practitioner, *qua* practitioner, of an art exclusively. Socrates claims that instead of discussing the shepherd "in the strict sense," Thrasymachus is speaking of the shepherd as an earner of wages or fees (345b5-346d7).

Other alleged inconsistencies have to do with the second major point of disagreement between Socrates and Thrasymachus (whether the just or the unjust life is more profitable) and will, therefore, not be discussed in detail.

* * *

In defending Thrasymachus against the above charges I require a bit of stage-setting:

(1) I shall assume that Thrasymachus and Socrates are in implicit agreement that just action requires for its occurrence some form of human society. Justice is treated throughout the dispute as a social phenomenon, demanding for its instantiation a context in which people act upon, interact with, and in various ways deal mutually with each other, and the same goes for injustice, of course. The examples of just actions appealed to in the text are all such as to involve normally some form of reciprocal action by those toward whom just action is directed: the mutual keeping of contracts and bargains, business deals, etc. (e.g., 343d1-e5).

(2) It is a truism to observe that genuine disagreements can occur only within a context of shared views. It is conceivable that Socrates and Thrasymachus share so few points of agreement that their whole dispute is a series of arguments at cross-purposes. Some commentators implicitly or explicitly presume that this is just what happens. I submit that there is no direct textual evidence for this conclusion. If it were true, it would constitute a serious indictment of *Plato*, since he nowhere has Socrates or any other character point this out. What I propose to do is to assume the widest possible range of shared views between the two

antagonists which is consonant with the text; that is, on each point relevant to the argument, if there is not some textual reference precluding it, I shall assume that Socrates and Thrasymachus are in general agreement. For example, since it is nowhere denied, I shall assume that on the whole they agree as to the general types of actions normally designated "just" and "unjust." Honoring contracts, paying debts, paying taxes, keeping bargains and so on would be some examples of actions which they could, and apparently do, agree are just actions and their opposites unjust. Their disagreement is not, or need not be, as to which types of acts are just, but rather as to the essential property which all just acts share.

(3) I shall also assume that Socrates and Thrasymachus both understand what is meant by an essential property, and that they are not arguing at cross-purposes in their attempt to arrive at the essential property shared by all just action, and only by just actions, which is designated by the universal term "justice." It might be objected that Thrasymachus is not seeking a definition of justice, but rather presents his conclusion baldly at the beginning. Nevertheless, he is willing to argue for the truth of his contention and to abide by the outcome of the argument. It is no logical error to state one's conclusion first and then present one's arguments.

* * *

Does Thrasymachus' formulation of his definition of justice in *Jb* contradict *Ja*? Perhaps, but not necessarily, and it is surely not poor scholarship to give Thrasymachus the benefit of the doubt. Cross and Woosley interpret *Ja* to mean that every time anyone acts justly the *consequences* of such acts will be in the form of some benefit or profit to the strong man.² But in *Ja* Thrasymachus does not refer to the consequences of just actions; he says that justice is nothing else than . . . etc. Why not take him at his word and assume that he is designating what he believes to be the essential feature of just action itself, rather than the consequences which accrue from performance of just action? If Cross and Woosley's interpretation is accepted, then Thrasymachus' account of justice can be a definition only if one interprets him as a subjectivist who is claiming that any action which results in benefit or profit to the stronger is, for that reason, just. The only reasonable alternative would be that

² R. C. Cross and A. D. Woosley, *Plato's Republic: A Philosophical Commentary* (London, 1964), cf. pp. 27, 37, 38 ff.

Thrasymachus is not giving a definition at all, but is simply presenting an empirical generalization. In either case Thrasymachus' view is much weaker than it need be.

But what does it mean to say of just action itself that it is advantageous to the stronger? Nothing very esoteric. I think Thrasymachus means that in the context of an ongoing, dynamic society, when two or more people (or groups) have dealings with one another, if one person (or group) acts justly toward another or others, the very act of justice renders the just agent vulnerable and susceptible to being unjustly exploited; that is, to being *taken advantage of*.

Consider Thrasymachus' examples (343d1-e5): when a just and an unjust man are business partners, the unjust partner always profits excessively in comparison with his just partner. Out of equal incomes, the just man pays more taxes than the unjust. Where there is money to be got the just man gets nothing, the unjust much. Why? Because by acting justly toward his fellow citizens, his ruler, or his business associates, the just man provides the opportunity for the unjust man to take advantage of him.

Suppose that two men, *A* and *B*, each have \$50 worth of goods and mutually agree to barter \$25 worth of goods with each other. If *A* performs his part of the bargain by delivering \$25 worth of goods to *B*, he would be acting justly with respect to *B*, and I think Socrates and Thrasymachus would agree. But in so doing, *A* thereby places himself in a position where he can be (unjustly or unfairly) *taken advantage of* by *B*. If *B* acts unjustly and reneges on his part of the bargain, he emerges, not twice as wealthy as *A*, but three times as wealthy.

In this example it should be noted, first, that *A*, in performing his part of the bargain, places himself at a tactical disadvantage (i.e., in a position to be taken advantage of) in his dealings with *B* *whether or not B decides to act justly*. Secondly, it is obviously true that *B* is not compelled to act unjustly toward *A*: *B* might willingly turn over to *A* a full \$25 worth of goods as agreed. If *B* does so, then *no one* benefits or profits in Cross and Woosley's sense, and thus, in this sense, neither *A*'s action nor *B*'s action could have been just.³ Thirdly, no mention of weak or strong men is made in the example. Thrasymachus does not say that a man is strong *because* just action is advantageous to him. On the present interpreta-

tion, just action always gives others, just and unjust, weak and strong, the opportunity to take unjust advantage of the just agent. Just action is said to be advantageous to the stronger, because the stronger is the one who will always seize the opportunity for exploitation afforded him by the just actions of others.

In contrasting my interpretation of Thrasymachus' view with that of Cross and Woosley, one should be careful not to confuse *causal* with *logical* relations. From their examples, Cross and Woosley seem to be taking Thrasymachus to mean that just action is defined in terms of its (causal) consequences. It might be argued that my interpretation does not differ significantly from theirs, for according to it, just action is also defined in terms of its (logical) consequences; that is, that it is a (logical) consequence of acting justly that one becomes vulnerable to unjust or unfair exploitation.

I am willing to accept this observation as true of my view, while denying that it trivializes the distinction between my interpretation and that of Cross and Woosley. The difference between them is as great as that, e.g., between the claim that the results of my throwing a stone was that it broke a window, on one hand, and the assertion that the results of my throwing a stone was that a stone was thrown, on the other hand. Cross and Woosley apparently assume that Thrasymachus has in mind the obvious fact that all actions have (causal) consequences, even if only in the minimal sense that something in the world must be different from what it would have been if the action had not been performed, and that he is asserting that all actions which have (causal) consequences of a certain general type are just actions. If I am correct, however, Thrasymachus is concerned instead with the (logical) consequences of just actions with respect to their agents, and with the types of responses by others which just actions render possible.

One must be careful not to confuse three concepts mentioned in Thrasymachus' account: the unjust man, the strong man, and the ruler. No one of these is identical in meaning to any other. Strong men are *always* unjust men, for Thrasymachus, but to be a strong man there are other requirements besides being unjust. To be a strong man one must first possess Thrasymachus' insight that justice is advantageous to the stronger. Second, one must possess the intelligence to practice injustice on a

³ Or, alternatively, one might say that, on Cross and Woosley's view, if both *A* and *B* perform as agreed, then *both* profit, entailing the inconsistent consequence that each is stronger than the other.

grand scale and get away with it. Finally, one needs the courage to engage in and pursue such a program. Thrasymachus points to examples of unjust men who lack either the wit or the courage or both to practice injustice on a grand scale (pickpockets, cutpurses, temple-breakers, etc.), and although he thinks of these as stronger than their counterparts of similar station among just men, they fall short of the ideal strong man (348d5-9). The just man, on the other hand, could never be a strong man, for he lacks at least one necessary quality: the understanding that justice is advantageous to the stronger.

The ruler "in the strict sense" is defined by Thrasymachus as one who always makes laws which are to the ruler's advantage (340e9-341a2). He is *defined*, be it noted, neither as an unjust, nor as a strong man.

Finally, the unjust man, on Thrasymachus' view, is simply a man who performs unjust actions, and as I pointed out above, since there seems to be no textual considerations precluding it, I am assuming that Thrasymachus and Socrates are in general agreement as to which actions unjust actions would be.

It is true that Thrasymachus argues that rulers of countries, whatever the form of government might be, are the real strong men. Yet, there is no textual reason to believe that for Thrasymachus the ruler is the strongest man because he is the ruler of a state. Rather, the strongest man in the state is most likely to be, or to become the ruler. He rises to the top naturally because he takes advantage of every opportunity to make an unjust profit and to further his own cause at the expense of others. Everyone and every group who deal with him justly are exploited by him for his own profit. It could hardly be the case, as some commentators have suggested, that the ruler is by definition the stronger, because Thrasymachus admits the possibility of there being a just ruler (343a). A just ruler would not be the strongest man in the state in Thrasymachus' sense of "strong": indeed he would not, in this sense, be strong at all. For Thrasymachus says that the just ruler loses on all counts: his business suffers through neglect, and he loses the respect of his friends and relatives because he will not grant them special privileges during his tenure of office.

It should be obvious by this time that if the above interpretation of Thrasymachus' position is accepted, there is no inconsistency between *Ja* and *Jb*. Just action, by its very nature, is advantageous

to the stronger, for the stronger by definition is one who takes advantage of all opportunities to benefit himself. Justice is the good of another, in that acts of justice afford others the opportunity to cheat and defraud the just agent. Injustice is advantageous to oneself, in that acts of injustice are those in which one takes unfair advantage of others. *Ja* does not imply, on my interpretation, that if the strong man acts justly it is to his own advantage. Rather, *Ja* entails that just action always creates opportunities for the unjust exploitation of just agents, which is repeated in slightly different words in *Jb*.

Let us consider in more detail what the life and character of the completely unjust ruler might be like, given Thrasymachus' view of the nature of justice and injustice. Some commentators seem to think that the only sort of man who could fulfill Thrasymachus' conception of such a ruler would be an absolute dictator who rules his country with an iron fist, stealing, killing, and imprisoning whenever his slightest whim is opposed. It is possible that Thrasymachus would agree that such a ruler would, or could, be a strong man in his sense of the term. But it should not be forgotten that he claims that the rulers in every state, no matter what the form of government, usually fit his conception of the strong man:

Well, every government lays down laws for its own advantage—a democracy democratic, a tyranny tyrannical laws, and so on. In laying down these laws they have made it plain that what is to their advantage is just. They punish him who departs from this as a law-breaker and an unjust man. And this, my good sir, is what I mean. In every city justice is the same. It is what is advantageous to the established government. But the established government is master and so sound reasoning gives the conclusion that the same thing is always just—namely, what is advantageous to the stronger (338e).

It follows that even in a country where the ruler is elected by majority vote, and where to retain his position as ruler, he must retain the esteem of the electorate, the ruler may still be, and usually is, the strongest man in the state. This means that such a ruler could be a complete villain, a man who believes that the unjust life is the best life for man, and that justice is advantageous to the stronger.

Bearing this in mind, let us consider a possible alternative to the view that Thrasymachus' ruler-type would have to be an iron-fisted dictator. Imagine the following rather extended hypothetical case:

Early in life a politically ambitious man named

Setarcos comes to believe that Thrasy-machus' position is entirely correct. Suppose that Setarcos is both intelligent and courageous. While taking advantage of every opportunity to profit unfairly and to advance his own fortunes at the expense of others, Setarcos is clever enough to maintain a public facade of honesty and integrity. He publicly proclaims his absolute faith in the view that the just life is the best and most profitable life for man. "Mutual trust is the cornerstone of society," he tells everyone, "and mutual trust is possible only when the citizens of a state deal justly with one another." His voice is loudest in condemnation of every unjust and dishonest act by an official in power. He ferrets out or, if need be, manufactures evidence of immorality and corruption against all those who oppose him or stand in the way of his rise to power. Eventually he campaigns for the highest office in the land on a reform ticket, pleading for a return to honesty, justice, and fair dealing in government. He pledges himself to the elimination of graft and corruption in high places. When he achieves power, he is too clever to perform an about-face and become a tyrant, perpetuating his rule by force. Rather, he decries the fraud, waste, ineptitude, and corruption which he claims is his legacy from the previous administration, and reiterates his belief that justice is the best policy. After all, he is a just man, and hasn't he become ruler of the land?

Once he becomes the ruler, Setarcos may be able, if he is clever enough and bold enough, to maintain his public facade of justice and honesty for a long time or even indefinitely, while remaining a thoroughly unjust man.

But is this at all a plausible account? If Thrasy-machus is correct, such a ruler, to the best of his ability, would always enact only those laws which are advantageous to himself. Do we have to imagine an entire citizenry so incredibly naive and innocent as to overlook this fact indefinitely? For if people were to notice that the laws of the land are of this nature, would this not immediately reveal Setarcos as a completely unjust man?

The answer to both questions is: Not at all. One of the strongest reasons for acceptance of my interpretation of Thrasy-machus' view is that it would entail the possibility that a completely just and a completely unjust ruler might enact exactly similar sets of laws! Thrasy-machus says that *justice*, not *injustice*, is advantageous to the stronger, so that an unjust ruler would be foolish to enact, e.g., repressive or discriminatory laws. He would want the citizens to act toward each other and toward

him as a ruler in ways which both Thrasy-machus and Socrates would agree are just. For Setarcos holds, with Thrasy-machus, that if the subjects believe that the just life is the most profitable, and thus are just and law-abiding in their dealings with one another and with their government, they will be most vulnerable to exploitation by the ruler. All states must collect taxes to finance legitimate governmental functions, for example. If the citizens realize this fact and willingly obey the tax laws of the state, they place themselves in a perfect position to be unjustly exploited by the ruler. It is often very hard for the subjects in a country to determine whether *all* of their tax money is being spent wisely and justly for the purposes for which it is collected. Furthermore, by convincing the people that they are serving their own interests by living completely just lives, Setarcos eliminates a major source of expense and anxiety which often plagues tyrants: he does not have to employ nearly as large an internal security force to preserve order, enforce his laws, and suppress possible rebellions.

It is interesting to note in passing that Thrasy-machus might well have argued that Socrates, who is known for his attempts to defend the just life as the best and most profitable life for man, is actually playing into the hands of the unjust ruler. Setarcos would want everyone in the state (except himself who knows better) to act justly, to live just lives, and to believe sincerely that in so doing they were serving their own best interests.

It is not merely because one sees clearly that acting justly renders the just agent vulnerable to exploitation that one can become a strong ruler of a state. He might lack the intelligence to effect a master plan of injustice and power-seeking, or else he might lack the courage for such a large-scale operation. In the context of the present hypothetical case, our unjust ruler, Setarcos, might be ranked as to his degree of "strength," in Thrasy-machus' sense, on the basis of how well he was able to convince the populace of the folly of living unjustly, especially in their dealings with the government of their state. If Setarcos were able to convince everyone in the state that he is a completely just man, that because he is just he is happy, that justice in general is most profitable to man as a way of life, while at the same time being able, covertly, to cheat and steal from the people systematically, then he would conform perfectly to Thrasy-machus' conception of the strong man.

If Thrasy-machus were to prove his case that the unjust life is the good life for man, would this entail

that Socrates, or you, or I, or anyone who understands what he is talking about *ought* to begin living the unjust life? This conclusion would at least not follow from such a proof. I find no textual evidence that Thrasymachus was advocating universal injustice as a way of life for everyone. In fact, the greater the ratio of just to unjust people in a state, the better life can be for the unjust few. Even if everyone were to begin to live completely unjust lives, however, this would not negate Thrasymachus' claim that justice is the advantage of the stronger, only no one would then be acting justly. At any rate, to describe a certain way of life as the best life for man is not equivalent to, nor does it entail that everyone should, or even can, live such a life. Thrasymachus could be implying no more than that Socrates and the others ought to wake up and realize that they are like sheep being kept for the profit of the unjust man, and that sheep-like, they assume naively that to be fleeced and slaughtered for mutton is better than wearing warm clothes and being well-fed.

It is important to understand that Thrasymachus does not argue that the ruler of a country *ought* to live an unjust life for the following reason: when Thrasymachus defines the "ruler in the strict sense" as the one who infallibly enacts laws which are advantageous to the ruler, he might be interpreted as meaning that it is logically impossible for a just man to be a ruler in the strict sense. If so, then of course he would be inconsistent, for he has admitted that a just man could be a ruler. But it would be a mistake to interpret Thrasymachus as claiming that when a ruler makes laws which give him the opportunity to exploit the people for his own profit, that he ought to exploit them, and that if he does not, he is not a ruler in the strict sense at all. On the contrary, the just man certainly could be a ruler in the strict sense, if we do not read into Thrasymachus' definition something which isn't there in the text. The laws which the just ruler makes, as pointed out above, might be exactly similar to the set of laws which an unjust ruler would make (even if neither made any legislative mistakes). Thus the just ruler's laws, like those of the unjust ruler, would afford him the opportunity of taking unfair advantage of the citizens when they act justly. Being just, however, he would not do so. But in failing to cheat the populace he is not failing to do something which he *ought* to do, if he is to be a ruler in the strict sense. On Thrasymachus' view, he is merely being stupid.

To sum up briefly, I interpret the dispute between

Socrates and Thrasymachus as a genuine disagreement on issues which are quite important, historically, for moral philosophy. It is a genuine disagreement, in the sense that it arises within a context of shared opinion, and on no major point are they arguing at cross-purposes. I view Socrates and Thrasymachus as being in broadly general agreement as to the practical content of the just and the unjust lives—that is, as to the types or kinds of actions which are correctly called "just" and "unjust." The dispute between them is not a simple semantic one regarding the correct moral designation of various types of actions. There is nothing in the text to suggest that they would not both consider just such actions as honoring contracts, paying taxes, obeying the law, giving honest measure, and so on.

I interpret Thrasymachus as claiming that just action is intrinsically disadvantageous to the performer because by its very nature it places the just agent in a vulnerable position with respect to those with whom he deals in practical life. By acting justly toward one's fellow man, *ipso facto*, one places oneself in a position to be unjustly taken advantage of. When Thrasymachus says that "justice is nothing else than that which is advantageous to the stronger" he is referring to this characteristic of justice, which he believes to be its essential property. By the "stronger," he means the person who sees justice and the just life for what he, Thrasymachus, believes it to be, and who has the intelligence and the courage to practice injustice on a grand scale. One who is truly a strong man in his sense would thus never voluntarily live a just life.

* * *

Let us now consider the major criticisms which Socrates offers to Thrasymachus' view:

If an unjust ruler makes a law which he mistakenly believes to be advantageous to himself, aren't the subjects acting justly when they obey this law? If so, then justice, in this case, would not be advantageous to the stronger. Cleitophon and Polemarchus immediately conclude that this criticism is devastating to Thrasymachus' stated position (339e9–340c4). If the essential quality of justice is that it is always advantageous to the stronger, then there could not be a case of justice which was not advantageous to the stronger. Socrates' criticism seems to them to offer a perfect counter-example to Thrasymachus' account of the nature of justice.

Cleitophon, however, believes that Thrasymachus has merely misstated his view, and suggests that "... by what is advantageous to the stronger he meant 'what the stronger thinks is to his advantage.' This is what the weaker must do, and this is his definition of justice (340b5-7)."

Cross and Woosley argue that Thrasymachus would have done well to accept Cleitophon's reformulation in order to escape inconsistency, "for while a ruler may make a mistake as to what actually *is* his interest he will hardly make a mistake as to what he *believes* to be his interest; and if it is right for subjects to do what the ruler believes to be in his interest, it will not matter that the ruler is mistaken in believing so."⁴

Sparshott⁵ disagrees with this suggestion, and I believe rightly so. If Thrasymachus *had* agreed with Cleitophon he would have been inconsistent. For he has claimed that justice is advantageous to the stronger, and Cleitophon's suggestion entails that justice is obedience to the law. If justice is nothing more nor less than obedience to the law, however, then the grounds or reasons for enacting the laws (the advantage of the stronger) drop out as irrelevant.

Thrasymachus, however, emphatically rejects Cleitophon's suggestion on the grounds that in the strict sense of ruler, a ruler never makes mistakes (340d1-341a2). Thrasymachus, of course, is not here claiming either that some rulers are, have been, or might be infallible, or that any ruler who is unjust will be infallible. His case does not depend upon the actual existence of an infallible ruler. He never says, e.g., that a doctor who prescribes the wrong medicine on a particular occasion is not really a doctor, nor that an accountant is not really an accountant if he sometimes makes mistakes in calculation. He simply says, in effect, that our grounds for calling a man a doctor or an accountant are not that he makes, or is capable of making, mistakes in the practice of his profession, and the same holds true for rulers.

Thrasymachus is, I suggest, doing something here which is quite common in contemporary moral philosophy: he is distinguishing a role or office from the man who holds the office or plays the role. If someone were to ask what a doctor is, and if we know of a particular doctor who has made a mistake in the diagnosis of a certain patient's illness, we would be responding in a misleading and

inappropriate way if we were to answer by saying that a doctor is one whose job it is to make mistaken diagnoses of people's illnesses.

There is still an important question at stake here, however: If a doctor prescribed a certain medicine on the basis of a mistaken diagnosis of a patient's illness, would the patient be acting as a patient should if he obeyed his doctor? By analogy, we might ask Thrasymachus: If the unjust ruler makes a law which he mistakenly believes to be advantageous to himself, would the subjects be acting justly if they obeyed this law?

We know, of course, that normally what the doctor prescribes is what a patient ought to do to treat his particular illness or affliction. It is the proper job of the doctor to prescribe the types of treatment which are appropriate to the particular ills and afflictions of his patients. Thus "doing what the doctor orders" is commonly accepted as roughly synonymous with "applying the appropriate treatment to the proper illness or affliction." But this does not mean that appropriate treatment for any given disease or affliction is correctly defined as "whatever the doctor prescribes." If I smashed my toe with a hammer and immediately consulted a doctor, his advice to have my leg cut off at the hip would not be acceptable (at least not if the smashed toe were all that was wrong with me).

Hence, in the case of Thrasymachus, we could say that in one sense, a vulgar or loose sense, a man would be acting justly if he obeys a law which a ruler mistakenly thinks to be to his own advantage, but in another more strict or absolute sense, he would not. One might today want to contest the legitimacy of such a distinction, but obviously Plato would not, since this is a characteristically Platonic way of arguing.

At any rate, Socrates accepts Thrasymachus' answer as satisfactory. Subsequently, he speaks only of the ruler in the strict sense, and indeed, as Nettleship⁶ points out, he thereafter identifies the practitioner, *qua* practitioner, of an art or profession with the art or profession.

Socrates next asks Thrasymachus whether the physician, *qua* physician, is a "money maker, an earner of fees, or a healer of the sick," and Thrasymachus, of course, says the last (341c2 ff.). And if a man is sick, it is obviously to his benefit to have the services of a healer available. Socrates then refers to other examples, each of which is designed

⁴ Cross and Woosley, *op. cit.*, p. 46.

⁵ F. E. Sparshott, "Socrates and Thrasymachus," *The Monist*, Vol. 50 (1966), pp. 424 ff.

⁶ Richard L. Nettleship, *Lectures on the Republic of Plato*, 2nd ed. (London, 1964), p. 30.

to show that it is the *subject matter* of an art, which is benefited by the perfect practicing of the art. Thrasymachus grudgingly agrees in each case. Thus in regard to the ruling of a state, which Thrasymachus has agreed is an art, it would seem by analogy that it is the subjects as "subject matter" of the art of ruling, who benefit from the perfect practice of this art, rather than the ruler who practices it.

It seems obvious in the text that Socrates believes that his argument from analogy of the other arts to the art of ruling has dealt Thrasymachus' account a death blow. It is at first difficult to see why he should think so. The principle question at issue has to do with the nature of justice. Even if it is true that the perfect practice of the art of ruling is beneficial⁷ to the subjects, why should this force Thrasymachus to give up his claim that justice is advantageous to the stronger? After all, he has not claimed that the ruler is a strong man, by definition, nor has he argued that in order to be a ruler, *qua* ruler, one must be unjust. Admittedly he would have to give ground on some points, but even so, why could he not rebut Socrates' argument successfully by saying, "All right, if a strong man becomes ruler, he won't act as a ruler should, strictly speaking, but he could still be a great strategist and an unparalleled grafter"?

I believe that the reason why Thrasymachus does not take this line is that Socrates (or perhaps Plato) has accepted a suppressed premiss which Thrasymachus fails to question: it is that the only reason why a practitioner of an art would fail to manifest perfectly the definition of that art in his practice is that he is ignorant. Either he does not know the definition, or else he doesn't know in every case which action instantiates the definition (although in the actual dispute, the latter kind of ignorance is seldom touched upon). When Thrasymachus first makes the vulgar-sense vs. strict-sense distinction, the examples he uses are those of a doctor who misdiagnoses a patient's illness and an accountant who makes a mistake in calculation. It is implicitly assumed that the examples referred to are examples of unintentional errors, and this kind of example is extended across the board to all the arts, including the art of ruling. Thus in the analogy of the arts, Socrates assumes without argument that the only deviations from perfection in ruling must be in the form of unintentional errors, due to ignorance of the true nature of ruling.

⁷ The term translated "advantageous" in *Ja* is the same as that translated "beneficial" in Socrates' analogy of the arts (Tō Xumpéron). Thus his claim could have been worded, "the perfect practice of an art is advantageous to its subject matter."

Because Thrasymachus accepts this suppressed premiss, he believes himself to be in the following dilemma: Insofar as a man is unjust, he is concerned only with self-aggrandizement; anyone who practices an art less than perfectly does so out of ignorance of the nature of perfect practice of that art; the perfect practice of the art of ruling is advantageous to the subjects, and not to the ruler. Therefore, an unjust ruler, to the extent that he is unjust, and in virtue of the fact that he is unjust, is ignorant of the nature of ruling. Hence a completely unjust ruler would be completely ignorant of the art of ruling. He would be directly analogous to the man who knows absolutely nothing about music or musical instruments, yet who attempts to attune a stringed instrument properly (cf. 349e).

Although Thrasymachus can see no way out of this dilemma, Socrates' conclusion seems to fly in the face of obvious facts. Thrasymachus believes that the vast majority of actual rulers are grossly unjust men, and, far from being ignorant of the nature of ruling, they seem to him much more knowledgeable than the ruler who has the opportunity to defraud his subjects on a grand scale, but deliberately refrains from doing so. Socrates' view seems to him incredibly naive. An analogy Thrasymachus might have used would be that of a gambling game between an honest man and a cheat, in which the cheat wins (by cheating) the money, lands, possessions, servants, slaves, titles, even the clothes on the back of the honest man. By analogy, Socrates' position would seem to imply that the honest man actually comes out ahead, because he plays the game as it should be played, and that every time the cheater cheats, he thereby merely reveals his ignorance of the game.

Surely Thrasymachus should have rejected Socrates' suppressed premiss. Nothing in Thrasymachus' account requires it. To accept this premiss would be to place in the same category of ignorance a young, inexperienced physician who misdiagnoses an illness, and/or mis-prescribes treatment, on one hand, and an experienced physician who prescribes removal of organs and tissue which he knows to be healthy (tonsils, appendices, etc.), to gain an undeserved fee, on the other hand.

Nevertheless, in the text Thrasymachus does not reject Socrates' suppressed premiss, but rather, in frustration, turns the full force of his scorn and derision upon him (343a1, ff.): Socrates needs a nurse to wipe his nose. He is such a child in

practical matters that he cannot see that shepherds and cattle herdsman tend sheep and cattle for their own or their masters' profit, rather than for the benefit of the sheep and cattle, and that the rulers of states treat their subjects like sheep, caring for them only to gain greater profit from them.

Socrates' answer to this attack (345b4, ff.) is to chide Thrasyarchus for failing to draw the conclusion which follows from the former argument (i.e., that the art of ruling is advantageous to the subjects and not to the ruler), and to maintain that the same kind of argument applies to the "art" of shepherding: a shepherd is correctly defined, not as an earner of wages or fees, but as one who cares for sheep. Thus again it is the subject matter of this art (the sheep) which benefits from excellent practice, rather than the practitioner.

It is interesting to note that Thrasyarchus never actually admits that Socrates has refuted his claim that justice is advantageous to the stronger. Thus we are left to decide for ourselves whether he really believed his view to be defeated, or, alternatively, whether he still believed his original position to be perfectly correct, even though he was unable at the time to see a mistake in Socrates' reasoning.

Does Socrates make such a mistake? Even if, purely for the sake of argument, we were to grant him his suppressed premiss (i.e., that any deviation from perfection in the practice of an art is unintentional and due entirely to ignorance of the true nature of the art), has he indeed shown that, in the appropriate sense, the art of ruling, strictly speaking, is beneficial (advantageous) to the subjects?

It is not unimportant that, in the dispute with Thrasyarchus, Socrates does not spell out the ways in which the perfect practice of the art of ruling would be advantageous to the subjects: he merely concludes that it *must* be so, because in the case of the other arts examined, the subject matter of the arts, rather than the artists, were the beneficiaries of excellent practice. Note also that it is absolutely vital to Socrates' case that the benefits derivable from the perfect practice of an art be seen as such by the subject matter, in the sense that the practice of the art must provide something needful, worthwhile, or desirable from the viewpoint of the subject matter. Granted this presents a real difficulty in those cases in which the subject matter is inanimate, as for example, in the manufacture of musical instruments: even here, by extension, one might think of the objective worth of, e.g., a violin being

enhanced by the skill of a master craftsman, whereas a hack might have used the same materials and made an inferior instrument.

In fact, Socrates does not focus attention on those arts whose subject matter is inanimate when he is trying to prove his point about the art of ruling. His primary example is that of the physician. A physician is defined as a healer of the sick. Healing the sick is an activity which is beneficial to the patients, in the sense of providing a service which is needful and desirable from the patients' point of view. It is this feature which Socrates believes has its analogue in all the other arts, including the art of ruling. But even though this may be true of medicine, it is a mistake to conclude that all arts, if practiced excellently, are desirable or needful from the point of view of the subject matter. This can easily be shown by appeal to other sorts of practices which would surely fit into Socrates' broadly general concept of an art.

The art of torture,⁸ for example, would surely fit Socrates' model. As an art, it can be practiced well or badly. The proficient torturer is the one who can keep his victims alive and in constantly increasing agony for the longest period of time. He is the one who never fails to extract the confession, or the recantation, or the oath of allegiance, or the suppressed information from unwilling victims. It follows that, just as in the case of the physician, one cannot correctly define a torturer, *qua* torturer, as an earner of wages or fees.

Thrasyarchus himself provides another example which, with a few hypothetical amplifications, could also serve to illustrate Socrates' mistake: Suppose it were the case that extremely fat sheep bring the best price on the market. Suppose, however, that very fat sheep suffer from shortness of breath, constant pain in the lungs, aching ankles, and continual nausea. As Socrates says, the shepherd, *qua* shepherd, is defined as one who cares for sheep, not as an earner of wages or fees. Yet in this case, no one could deny that the best shepherd would be the one who was able to bring the fattest sheep to market (i.e., this is what "caring for the sheep" would consist of, at least in part).

The above are only two of many possible examples which could be given to illustrate the point that the perfect practicing of an art may not be such as to fulfill the needs of, improve, or in other ways be desirable from the point of view of the subject matter.

Socrates speaks of the function of an art as, in

⁸ This example was first suggested to me by T. G. Smith.

one way or another, alleviating the "defects" of its subject matter (342a2, ff.). The art of medicine has come into being because the human body is defective. But this sort of language unjustifiably prejudices the case. In what sense is the subject matter of the art of torture "defective"? Only in the sense that various forms of external stimuli are capable of causing people pain. The art of torture, then, would never have come into being if people were unable to feel pain and anguish.

If Socrates' discussion of the function and origins of the various arts are purged of such question-begging language, what remains? What conclusion can one draw from his analogy of the arts? I suggest, merely the value-neutral point that for any art or artist to exist, there must be an appropriate subject matter on which the art can be practiced. For there to be physicians and an art of medicine, there must exist patients on whom it can be practiced; i.e., people with bodies that are not invulnerable to disease, accident, or infirmity. For the art of shepherding to exist, there must be sheep to care for. And for there to be rulers, there must be subjects to rule. Whether the subject matter "benefits" from the proficient practicing of an art, in the sense of having something needful or desirable provided for it or done to it (from the standpoint of the subject matter itself, that is), is contingent upon the type of art involved. It is certainly not true of all the arts.

Socrates' argument that the perfect practice of the art of ruling is necessarily advantageous to the subjects thus fails completely to refute Thrasymachus. All that one is justified in concluding from his argument is that there must be subjects on which to practice this art, or else neither art nor artist could exist.

One can now also draw the conclusion that even if Socrates is granted his highly questionable premiss that failure to conform perfectly to the definition of an art entails ignorance on the part of the artist, it remains possible that a ruler might be unjust and still be a ruler in the strict sense.

If the textual Thrasymachus had seen the error in Socrates' analogy of the arts, he could have turned the tables on him by pointing out that Socrates is not dealing with the ruler, *qua* ruler, but rather with the ruler, *qua* just man; which, of course, is precisely what Socrates is doing. On Thrasymachus' view, one is a ruler in the strict sense of this term if he always makes laws which are advantageous to himself. As we have seen, it would be a mark of ineptitude on the part of the

ruler if these laws were other than what both Socrates and Thrasymachus would agree were just laws. If a ruler were in this sense infallible, he would be a just ruler in the strict sense of ruler if he did not take unfair advantage of the opportunities for exploitation which are afforded him by the law-abiding acts of his subjects; he would be an unjust ruler in the strict sense of ruler if he did cheat and defraud his subjects in their just dealings with him.

* * *

A transition in Socrates' line of argument becomes evident at this point. Socrates obviously believes that the appeal to the analogy of the arts has refuted Thrasymachus' claim that justice is the advantage of the stronger, and he then turns to the question whether the just life or the unjust life is more profitable, a very complex problem, but separate and distinct from the question of the essential nature of justice, and thus beyond the scope of the present inquiry.

I believe that Plato views Thrasymachus' account of the nature of justice as plausible and persuasive, and as one which, as far as it goes, is accurate. Justice as a way of life is a social phenomenon, as Socrates and Thrasymachus implicitly agree. It requires the mutual interaction of the members of a society or a social group for its occurrence. And it is surely true that just action, in the absence of any legal guarantees or collateral held, does place the just agent in a position to be unfairly exploited. And it further strengthens Thrasymachus' case to point out that in most organized societies there *are* such guarantees. For this is an admission that the vulnerability involved in acting justly has to be compensated for by the imposition on society of a system of laws, police, courts, and prisons to protect the just from the unjust.

To counter Thrasymachus' insight, Socrates could have attempted to show that a just man need not be so naive as to believe that all the people with whom he deals will be just. He could have examined the many ways, both legal and social, by which individuals and societies try to protect themselves against the depredations of unjust men. Indeed, this seems to be the sort of line pursued by Glaucon and Adeimantus in Book II. Socrates does not take this line, however, but rather attempts to defend the much stronger claim that justice is intrinsically advantageous as a way of life, in that it affords for those who choose it rewards far greater

V. ON A CERTAIN FORM OF PHILOSOPHICAL ARGUMENT

HENRY E. KYBURG, JR.

I

THERE is a certain form of philosophical argument, generally beginning with the words, "We have no reason to suppose that . . .," which depends for its force on the non-existence of certain inductive relations. Despite the fact that this form of argument is used without apology by philosophers of unquestioned eminence, and despite the fact that it can be used to support ontological, epistemological, and metaphysical conclusions of striking novelty, I think that such arguments are fallacious. In what follows I shall present several such arguments and exhibit the common fallacy in each one; I shall then attempt to characterize both the argument form and its fallacy abstractly, and to propose therapy.

II

The first illustration comes from C. D. Broad's long essay, "The Principles of Problematic Induction," published in 1928.¹ Broad discusses various matters concerning counters in bags, subject to the usual assumptions about stirring the counters so that every sample has the same probability of being drawn. (Incidentally he discusses there the alternatives of assigning equal probabilities to structure descriptions and to state descriptions; he concludes that only the assignment of equal probabilities to state descriptions—corresponding to Carnap's *c-dagger*—is philosophically defensible.) Having dealt with the balls-in-the-bag model, he then considers the relation between the model and the real problem of drawing inferences about the world from a sample of it. He says, "We come now to the difference which is most serious. In the case of the counters in the bag, we assumed that, at any drawing, any counter then in the bag was equally likely to be drawn, and this was an *essential premiss* of the inductive argument. Now, if

the bag be not too large, and does not have pockets in it, and the counters be well mixed, this assumption seems to be justified; but it most certainly is not justified in applying induction to nature. It breaks down for two reasons. (i) Spatially, only a very limited range is open to our observation. There may be swans on other planets, and, if there are, none of them could possibly have been included among our data. (ii) Similar remarks apply to time. Obviously the swans that could be observed up to a given date could not include any swans that began after that date, and it is equally certain that our observations . . . do not include swans that existed more than a few thousand years ago.² In short Broad's conclusion is that the universe is too large, has too many pockets, and is insufficiently stirred up to provide us with what he regards as an *essential premiss* of inductive argument.

The fallacy is that what Broad has exhibited as an essential premiss in one form of argument need not at all be an essential premiss in some other form of inductive argument. Surely one would not argue that because "All men are mortal" is an essential premiss in one form of argument whose conclusion is that Socrates is mortal, it must also be an essential premiss in every argument whose conclusion is that something is mortal.

So far as Broad has argued, there may be any number of alternative forms of inductive argument. His argument is therefore inconclusive. I shall exhibit one such alternative, which will show that, in addition, his conclusion is wrong. Consider a bag of counters. Draw (with or without replacement—it makes little difference) a set of n counters. It is a logical truth that practically all sets of n counters selected from the counters in the bag will be representative of the proportion of (say) red counters in the bag. "Practically all" and "representative" can be made as precise as we wish with epsilons and deltas. We have the two facts, one observational, and one logical:

¹ C. D. Broad, "The Principles of Problematic Induction," *Induction, Probability and Causation* (Dordrecht, 1968), pp. 86–126.

² *Ibid.*, p. 20.

- (1) Our sample, S , is a member of the set P_n of all sets of n counters drawn from the bag;
- (2) Practically all the members of P_n (the set of all sets of n counters drawn from the bag) are representative of the proportion of red counters in the bag; let us refer to the latter property as *Rep*.

These two facts, together with the epistemological fact,

- (3) Relative to our body of knowledge, S is a random member of P_n , with respect to possessing the property *Rep*,

yield the conclusion:

- (4) Relative to our body of knowledge, it is highly probable that the sample S is representative of the proportion of red counters in the bag.

The epistemological fact (3) requires spelling out. I have provided relatively detailed analyses of it in a number of places.³ Here it suffices to reflect on the way in which people use the probability concept in insurance, in quality control, in games of chance such as stud poker and horse-racing in which probabilities change in response to changes in background information. In all of these cases, we say that, relative to our body of knowledge, an object A is a random member of a set B , with respect to exhibiting a property C , when there is no subset B^* of B such that (a) we know that A belongs to B^* , and (b) we know the proportion of members of B^* that have the property C , and that this proportion differs from the proportion of members of B that have the property C . Let us put the matter negatively: we have reason to deny that A is a random member of B , with respect to exhibiting the property C , when two conditions are satisfied:

- (a) we know that A belongs to B^* , where B^* is a subset of B ; and
- (b) we know the proportion of members of B^* that have the property C and that this proportion differs from the proportion of members of B that have the property C .

Now let us observe that the inductive argument presented as (1), (2), (3), and (4) requires no premiss concerning mixing or the absence of pockets in the bag. Suppose that we take the n counters all from the top of the bag. Is this sample

a random member of P_n , with respect to the property *Rep*, relative to what we know? The answer depends on what we know. If we know that the counters were put into the bag in batches, one color at a time, and that they have not been stirred up, then S will not be a random member of P_n . In that case, there is a subset of P_n , say CP_n , satisfying conditions (a) and (b) above. CP_n is the set of sets of n counters drawn from the same part of the bag; S belongs to CP_n . And in virtue of our knowledge about how the bag was loaded, we know that the proportion of samples drawn from the same part of the bag that are representative is very low. Lacking knowledge of the proportion of representative samples among the members of CP_n , clause (b) will not be satisfied, and the fact that we know our sample belongs to CP_n as well as to P_n will not prevent it from being a random member of P_n with respect to exhibiting the property *Rep*.

We have now constructed an argument concerning counters in a bag which does not require any premiss about mixing, small size, absence of pockets, or the like. The swan argument is analogous.

- (1s) Our sample of swans S_s , is a member of the set P_{s_n} of all sets of n swans in the universe.
- (2s) Practically all the members of P_{s_n} are representative of the proportion of white swans.
- (3s) There is no subset CP_{s_n} of P_{s_n} such that
 - (3sa) we know that our sample S_s belongs to CP_{s_n} , and
 - (3sb) we know the proportion of representative samples in CP_{s_n} , and that that proportion differs from the proportion in P_{s_n} .

Therefore:

- (4s) Relative to our body of knowledge, it is highly probable that the sample S_s is representative of the proportion of white swans; i.e., it is highly probable that practically all swans are white.

It is clear how Broad would reply to this argument: he would deny the third premiss, that our sample S_s is a random member of P_{s_n} with respect to the property of being representative. He would say: "But we know that our sample belongs to a special—a very special—subclass of P_{s_n} , namely,

³ H. E. Kyburg, Jr., *Probability and the Logic of Rational Belief* (Middletown, Connecticut, 1961); "Probability and Randomness," *Theoria*, vol. 29 (1963), pp. 27-55.

the subclass consisting of all those sets of n swans drawn from (observed in) a relatively small region of space and time." This indeed does qualify as the set CP_{s_n} , and does satisfy (3sa). But in order that this should provide grounds for our rejection of premiss (3s), we must also have reason to accept (3sb)—i.e., we must have *reason to believe* that the proportion of representative samples in CP_{s_n} is relatively low. Broad has given us no such reasons. In point of fact we have such reasons in biological theory, but since they depend on positive scientific inference far more sophisticated than those being discussed by Broad, they are irrelevant to the argument at hand.

III

The second illustration comes from W. T. Stace's well known 1934 essay, "The Refutation of Realism."⁴ There he is concerned to defend the thesis that

There is absolutely no reason for asserting that these nonmental, or physical entities [sense objects] ever exist except when they are being experienced, and the proposition that they do so exist is utterly groundless and gratuitous, and one which ought not to be believed.⁵

Consider, he says, a piece of paper.

I am at this moment experiencing it, and at this moment it exists, but how can I know that it existed last night in my desk, when, so far as I know, no mind was experiencing it?⁶

He argues that we cannot know this by perception; that we cannot know it by deductive inference; and that we cannot know it by inductive inference.

It is clear . . . that any supposed reasoning could not be inductive. Inductive reasoning proceeds always upon the basis that what has been found in certain observed cases to be true will be so also in unobserved cases. But there is no single case in which it has been observed to be true that an unexperienced object continues to exist when it is not being experienced; for, by hypothesis, its existence when it is not being experienced cannot be observed . . . there is not a single case of an unexperienced existence having been observed on which could be based the generalization that entities continue to exist when no one is experiencing them.⁷

We find here the same sort of argument we found in Broad, this time leading to an ontological conclusion. Inductive arguments are characterized (they are arguments leading from the observed to the unobserved), and it is asserted that we have no inductive reason to believe a certain conclusion (that unobserved entities exist) since an *essential premiss* is lacking (that we have observed a number of unexperienced entities to exist). Again we can construct an inductive argument that yields the conclusion alleged to be unreachable. Let us consider a certain period of time, say 24 hours, and construe it as an ordered set T of small intervals, dt_1, dt_2, \dots, dt_m . Let us suppose that during n intervals comprising a certain subset of T , we observe Stace's piece of paper. Throughout the intervals during which we observe the piece of paper it exists. If bare existence seems queer, we can substitute spatio-temporal existence, or the possession of such properties as whiteness and rectangularity and flatness. Surely if something possesses such properties as these, it also possesses existence in Stace's sense. Let us say that a time interval dt_i has the property P if the piece of paper exists during it. We have the following argument:

- (1p) The sample of n time intervals during which we observe the piece of paper, Sp , is a member of the set TP_n of all sets of n time intervals belonging to T .
- (2p) Practically all members of TP_n are representative of the proportion of time intervals in T which have the property P , i.e., during which the paper exists.
- (3p) Relative to what we know, Sp is a random member of TP_n with respect to being representative of the proportion of time intervals in T which have the property P .

Therefore,

- (4p) It is highly probable, relative to what we know, that the piece of paper continued to exist during practically all of the intervals that comprise T .

What has Stace to say to this argument? Clearly he would deny (3p); the set of intervals Sp , he would say, is not just any old set of n intervals, but a very special one, for during each of those intervals we were experiencing the piece of paper. Thus if we

⁴ W. T. Stace, "The Refutation of Realism," *Mind*, vol. 43 (1934), pp. 145-155.

⁵ *Ibid.*, p. 365.

⁶ *Ibid.*, p. 366.

⁷ *Ibid.*, p. 367.

let CTP_n be the set of sets of n time intervals during all of which we were experiencing the piece of paper, we must accept the statement:

(3pa) Our sample Sp belongs to CTP_n .

But for this knowledge to serve as grounds for rejecting (3p), we must also have knowledge concerning the proportion of members of CTP_n that are representative in the appropriate sense. And obviously, we have no such knowledge; we do not have any grounds at all for supposing that the proportion of members of CTP_n which are representative of the proportion of elements of T that have the property P is any less than the proportion of members of TP_n itself which are representative of the proportion of members of T that have the property P .

Like Broad, Stace has made a gratuitous assumption about the form of inductive argument, and has used the fact that he cannot find one of the premisses required by an argument of that arbitrary form as grounds for accepting an extraordinary philosophical conclusion.

IV

Another instance of this same technique of argument—this time in epistemology—can be found in Chisholm's elegant little book, *Theory of Knowledge*.⁸ Chisholm, in the chapter "The Indirectly Evident," is seeking to uncover some useful relation between propositions of the sort that he wishes to regard as directly evident (such autobiographical propositions as "I take something to be a cat on the roof," "I seem to recall that it was there before," etc.) and propositions of the sort that he wishes to regard as indirectly evident. Among those "truths of fact" that are known but not directly evident, and thus among those truths that may be said to be indirectly evident, is "whatever we know about 'external objects,' . . ."⁹ including, for example, the proposition that there is a cat on the roof. Like Stace, he argues plausibly that the relation between these two sorts of propositions cannot be deductive; like Stace he argues implausibly that the relation cannot be inductive either. Unlike Stace, he considers not one, but several kinds of inductive argument. But he too is quick to leap to the conclusion that because he hasn't come up with

an appropriate inductive argument, none is possible. He therefore concludes that we must

. . . say that there are principles of evidence, other than the principles of induction and deduction, which will tell us, for example, under what conditions the state we have called "thinking that one perceives" will confer evidence or confer reasonableness, upon propositions about external things.¹⁰

This is a remarkable conclusion: we must accept, *a priori*, brand new principles, above and beyond the ones we always thought (without much thinking) to be sufficient for human inference: namely, inductive principles and deductive principles. Something new! And Chisholm goes on, ingeniously, to indicate what the character of this something new must be. With these new principles we are not directly concerned; we are concerned only with the argument which is alleged to establish their necessity.

Chisholm has no trouble in showing that deductive arguments cannot lead us from propositions that are directly evident to propositions that are indirectly evident, and he has no trouble, either, showing that no enumerative inductive argument can lead from statements like "I take something to be a cat on the roof" to a statement like "There is a cat on the roof." But he also considers the following sort of inductive argument for a hypothesis:

The premisses tell us, first, some of the things that would be true if the hypothesis were true, and secondly, that some of these things are true. . . . Thus, we might appeal to a generalization telling us some of the things that would happen if a cat were on the roof. We then perform a test or experiment to see whether these things are happening; if we find that they are, we argue that our hypothesis has been confirmed.¹¹

Chisholm considers the following specific argument:

If there is a cat on the roof and if I stand in the garden and look toward the roof, then I will see something that I will take to be a cat.

I am standing in the garden and looking toward the roof. I see something I take to be a cat.

Therefore, in all probability, a cat is on the roof.

He argues that this argument requires a premiss such as the first premiss above, which is a factual, synthetic statement telling us what would happen if

⁸ R. Chisholm, *Theory of Knowledge* (Englewood Cliffs, New Jersey, 1966).

⁹ *Ibid.*, p. 38.

¹⁰ *Ibid.*, p. 62.

¹¹ *Ibid.*, p. 4c.

a cat were on the roof, and that any argument of the same sort would have the same shortcoming.

This claim, that a general synthetic premiss is required, is questionable. It is surely not true that in an application of the hypothetico-deductive method the premiss which tells us "some of the things that would be true *if* the hypothesis were true" is inevitably a synthetic premiss. On the contrary, this premiss is ordinarily construed as a logically true conditional, whose antecedent is the conjunction of the hypothesis under test (say *H*) and certain boundary conditions (say *B*), and whose consequent is some observable state of affairs, (say *O*). It is true that in ordinary scientific discourse, as in detective stories, a number of the boundary conditions are left implicit, so that the major conditional premiss may not appear to be a logical truth. In keeping with this classical whodunit tradition, Chisholm has suppressed a few parts of the boundary conditions. Filling out these boundary conditions, we would get something like:

If there is a cat on the roof (*H*) and if I am in a normal condition and standing in the garden looking toward the roof under normal illumination, and attending to what I see (*B*) then I shall see something I take to be a cat (*O*).

If this seems less than logically true, regard it as the instantiation of:

(*P*) If there is a cat on the roof, then for any *x*, if *x* is a normal human observer in normal state, standing in the garden looking toward the roof under normal illumination, attending to what he sees, then he will see something he takes to be a cat.

Surely if the consequent of this conditional is false, we will regard the antecedent as false. There are a lot of proviso's regarding normality, attention, and the like; but let us not worry too much about them, for this is not quite the premiss we are looking for anyway. The reason is not far to seek: in using (*P*) to construct a hypothetico-deductive argument, we need in addition premisses of the sort: I am a normal human observer in a normal state standing in the garden looking toward the roof under normal illumination attending to what I see, etc. Some parts of this statement may be directly evident to me, but it is clear that others (such as "I am standing in the garden") are at best indirectly evident. Thus (*P*), though plausibly regarded as analytic, will not provide a framework for the

inference from the directly evident alone to the indirectly evident.

Part of the difficulty that arises in Chisholm's example is due to the fact that in his argument the assertion that there is a cat on the roof is, relative to the assertion that I take something to be a cat on the roof, not only a generalization, but a theoretical construction. That is, if for some obscure reason autobiographical statements are to be our data, then physical object statements must be regarded as relatively theoretical. Now it is well known that in general a theoretical statement cannot be tested in isolation; it can only be tested in a framework which includes boundary conditions, other theoretical statements, etc. The whole framework, and therefore also the theoretical statement under investigation, is supported inductively when its consequences are found to be supported. No doubt the inductive relation between observational statements and theoretical statements within frameworks is complicated; but perhaps we can reconstruct our highly artificial argument for the existence of a cat on the roof in the following way:

(*P**) It is an analytic consequence of there being a cat on the roof, that practically anybody, who takes himself to have good vision, to be looking at the roof under appropriate conditions of illumination and perspective, etc., will take something to be a cat on the roof.

A further consequence of there being a cat on the roof during Δt —though it may in part depend on independently confirmed parts of the theoretical structure—is:

(*C*) Practically always, when I take myself to be looking toward the roof during Δt , under what I take to be appropriate conditions of illumination and perspective and health, I shall take something to be a cat on the roof.

Here, finally, we have precisely the sort of statement that may be inductively confirmed by the techniques suggested previously. What indeed could prevent the set of times during which I do take myself to be looking at the roof under what I take to be appropriate conditions of illumination and perspective, from being a random sample of the set of times during which I shall take something to be a cat on the roof? The only answer that has any plausibility at all is that this set of times is distinguished by the fact that they are not only times

when I *could* have taken myself to be looking at the roof . . . , but also times when in actual point of fact I *do* take myself to be looking at the roof. Again, however, this cuts no ice unless we have reason to accept a statistical hypothesis to the effect that the proportion of representative samples among this subset is less than the corresponding proportion among the members of the original set. And we have, of course, no reason to accept any such statistical hypothesis.

Although it is not strictly to the point here, I might mention that I regard this epistemological reconstruction of perceptual judgments as quite wrong-headed. This reconstruction is based on a correct observation, that sometimes perceptual judgments of the sort, "I see a cat on the roof," are mistaken; but then the (to my mind) improper question is raised: "Since these judgments may be mistaken, they must be reached by probabilistic argument from judgments that cannot be mistaken. What is the nature of this argument?" Chisholm argues that it is neither inductive nor deductive, but hinges on what he calls principles of evidence. I have argued that we can perfectly well take the argument to be inductive, by construing the physical object statement "the cat is on the roof" as part of a theory which serves to explain relations among the autobiographical judgments that Chisholm wants to take as fundamental. A more plausible reconstruction of the epistemological situation is the following: Sometimes perceptual judgments are mistaken. How do we know when such a judgment is mistaken? When it comes into conflict with our body of reasonable beliefs about the world, which comprises judgments arrived at inductively, as well as perceptual judgments. Thus when the cat that I see on the roof flies to the top of a tree, I decide to retract my perceptual judgment that there was a cat on the roof, rather than either my perceptual judgment that I saw it fly into a tree, or my scientific generalization that cats don't fly. Within this framework the general defense of a perceptual judgment would be the following:

- (1f) This judgment is a member of the class of perceptual judgments.
- (2f) Practically all the members of this class are veridical; or, only a small portion of this class ever comes to be rejected in virtue of coming into conflict with other judgments.

- (3f) This judgment is a random member of the class of perceptual judgments, with respect to being veridical, or with respect to the property of not being rejected in virtue of coming into conflict with other judgments, relative to our body of knowledge.

Therefore,

- (4f) Relative to our body of knowledge, the probability is very high that this judgment is veridical, or that this judgment will not be rejected in virtue of coming into conflict with other judgments.

(I leave it an open question whether "being veridical" and "not being rejected by virtue of coming into conflict with other judgments" are the same thing. I suspect they are.)

Despite the fact that these appear to be two very different ways of reconstructing perceptual judgments, they stem from rather similar considerations. Chisholm claims that "evidence" is just as obscure as theoretical inference in general; I claim that evidence is just as straight-forward as theoretical inference in general. Like Chisholm, I feel that "... a proposition should be treated as innocent until proven guilty."¹² But Chisholm believes in guilt by association: if a certain sort of judgment admits of error, all such judgments are infected with serious doubt; while I believe in requiring proof of guilt in specific cases: although a certain proposition might be in error, we must have *reason* in order to doubt it seriously, i.e., in order not to accept it. Chisholm argues that among the truths of fact that are known but not directly evident, and that may be said to be indirectly evident, are "whatever we know about 'external objects' . . ."¹³ I should say on the contrary that there is nothing we know about more directly than external objects, and few things in ordinary life that we know about more indirectly than our internal states. But this is a subject for another essay. What is important here is only Chisholm's claim that new and different principles, other than induction and deduction, are required to pass from autobiography to natural history. I am arguing that induction will do.

V

Here is a final short example; I have heard similar arguments many times, but I know of no

¹² R. Chisholm, *Perceiving* (Ithaca, New York, 1957), p. 9.

¹³ *Theory of Knowledge, op. cit.*, p. 38.

place where it appears conveniently and succinctly put. The argument is as follows: we have no reason to suppose that the differential equations of physics (for example) hold all the time. We have, to be sure, excellent inductive reasons for supposing that when anything happens, it happens in conformity to these equations. This is not scepticism about induction. But the observational evidence would all turn out just the same way if there were vast periods of time during which nothing whatever happened, the planets hung fixed in the sky, people sat still at their typewriters, etc. Therefore we have no reason to suppose that the equations of physics hold all the time, but only that they hold while something is happening. Note the difference between this problem and Stace's; Stace says we have no reason to suppose the universe is behaving itself when we aren't looking at it; this argument says that so long as anything is happening, we have reason to suppose it happens in conformity to our physics, but that we have no reason to suppose that there aren't long periods of time during which nothing happens at all. But the answer is essentially the same:

- (1t) The set of time-intervals t_u during which we observe the universe is a member of the set T_n of all equi-numerous sets of time intervals.
- (2t) Practically all the members of T_n are representative of the proportion of the time during which the universe conforms to the laws of motion, conservation of momentum, etc.
- (3t) t_u is a random member of T_n , with respect to this property; i.e., there is no subset of T_n —neither the subset consisting of sets of n observed time intervals, nor the subset consisting of sets of n time intervals during which *something was happening*—of which we know (a) that t_u belongs to it, and (b) that the proportion of representative elements of it is less than the proportion of representative elements among T_n in general.

Therefore,

- (4t) It is highly probable that the differential equations of physics hold *all* the time, and not merely when someone happens to be looking, and not merely when something happens to be happening.

VI

The general form of the conclusion of all these arguments is the same: There are no premisses that will allow us to infer C . The form of the argument is the same in each case: such and such premisses P would allow us to infer C ; we cannot accept such and such premisses; therefore we cannot accept C . Sometimes (as in the case of Stace's argument) a further step strengthens the conclusion into: therefore we must accept not- C . When it comes to deductive arguments, one can sometimes do this: one can show that the denial of C is compatible with every set of premisses of a certain form, and this will show that no premisses of the required form can yield the conclusion C . For inductive logic, no such techniques are available, not only because inductive logic has not as yet been formalized, but because the whole essence of inductive logic lies in the fact that the denial of the inductive conclusion is (deductively) compatible with the premisses from which it is inductively inferred. It is therefore, *prima facie*, preposterous to pretend to show by means of a survey of all possible inductive premisses, that there are no inductive premisses that would confer probability on a given conclusion C . Yet this is just what Broad, Stace, and Chisholm have claimed to have done.

The discussions in question often focus on the propriety of a certain sort of statistical argument. Let us consider it in the abstract. We have a set of entities V : they may be time intervals, locations, events, objects. A certain subset of V has been observed in the past: let us call it O_p . Every member of O_p (or a certain proportion of members of O_p) has a certain property that can be expressed by an open sentence, "... x ...". Let O_e be the class of those members of V which it is physically possible for us to observe, O_{fp} those members of V that it is physically impossible for us to observe, and O_l those members of V that it is logically impossible for us to observe. In general, let $[X]_n$ consist of all those subsets of X that contain n members.

$$\begin{aligned} \text{We have } O_p &\in [V]_n \\ O_p &\in [O_p]_n \\ O_p &\in [O_e]_n \\ O_p &\in [V - O_{fp}]_n \\ O_p &\in [V - O_l]_n \end{aligned}$$

Let Rep be the property of reflecting (within some interval delta) the proportion in V of entities having the property expressed by the open sentence "... x ...". We have:

The proportion of members of $[V]_n$ having *Rep* is very high.

(This is true on logical grounds alone.)

We consider the argument:

- (1) $O_p \in [V]_n$
- (2) the proportion of members of $[V]_n$ having *Rep* is high.
- (3) O_p is a random member of $[V]_n$ with respect to *Rep*, relative to our body of knowledge.
- (4) the probability is high that O_p has *Rep*.

The only premiss in this argument that can plausibly be attacked is (3). As we observed earlier, the denial of premiss (3) amounts to the assertion that there is a subset of $[V]_n$, X , such that we have reason to believe that O_p belongs to X , and such that we know the proportion of X 's that have the property *Rep*, and that this proportion differs from (is less than) the proportion of *Rep* individuals in $[V]_n$. The plausible candidates for X are listed above. In the case of none of them do we have any *a priori* grounds whatsoever for supposing that the relative frequency of *Rep* is any different than among the members of $[V]_n$ itself.

There are certain implausible candidates: consider the set D , consisting of just those elements of $[V]_n$ to which *Rep* does not apply. Form the union of the unit class of O_p and D .

- (a) this is a subset of $[V]_n$;
- (b) O_p belongs to it;
- (c) the frequency of *Rep* is much lower (it may even be 0) in the union of the unit class of O_p and D than it is in $[V]_n$.

Does this prove that after all we cannot make inferences from an observed sample to the whole universe? Not at all; for the same kind of consideration would interfere with the most ordinary probability arguments. Suppose we know of a coin that it is well balanced. The probability of heads on the next toss is a half, because half the tosses result in heads, and the next toss is a random member of the set of tosses. But consider the set of future tosses that result in heads; form the union of that set with the unit set of the next toss; and it is clear that the coin belongs to a subset of the set of all tosses in which the relative frequency of heads is very high! So it would not, after all, be a random member of the set of all tosses, with respect to the property of landing heads, and the probability would not be a half that it will result in heads. The only way of having any probability statements at all

is to provide some sort of restrictions on the sets that may be allowed to prevent assertions of randomness. That is an open problem. I mention it here only to point out that it is as much of a problem for perfectly ordinary probability statements as it is for the more complex ones that are discussed in inductive logic, in metaphysics, in epistemology, and the like.

The objection may be raised that it is the very conception of randomness that I have adopted that is at issue; perhaps that conception of randomness begs the question by putting the *onus probandi* on Broad, Stace, and Chisholm? There is a flat answer: namely, that what they asserted was that there was no argument leading to such and such a conclusion; I have offered such an argument. There is an end of the matter. My speculative reconstruction of what they might say in reply to my argument is sheer generosity on my part, and represents what I take to be plausible in their claims concerning inductive arguments. It is plausible, but I do not take it to be persuasive enough to support their claim, because I take it that an assertion about randomness is essentially an epistemological assertion and is the sort of assertion that should be regarded as innocent until proven guilty (to use Chisholm's phrase again). Given such an interpretation of inductive logic, the arguments I have offered are not only perfectly relevant arguments of the sort they claim not to exist at all, but are, until positive grounds are offered for rejecting the assertions of randomness on which the arguments are based, perfectly valid inductive arguments. They can be refuted in a perfectly definite constructive way, by finding a set of a certain sort. Broad, Stace, and Chisholm seem to claim that such a set can always be found; but in none of the specific instances they talk about have they exhibited it.

Furthermore, the question of randomness is certainly not what is directly at issue here. Neither Chisholm nor Stace talk about randomness, and Broad, though he talks about randomness, is quite unclear in what he says about it. He wants his samples to have the same probability of being chosen as any other sample, and he wants to use "probability" in a logical sense; but he does not stipulate to what body of statements these probabilities are to be relative. A more common notion of randomness which might be thought to undermine some of the arguments I have offered, and which might be the notion that Chisholm and Stace have in mind, is a statistical one. In this sense

a sample is random if and only if it is selected by a method such that if that method were to be employed many times in selecting a sample, it would select each distinct sample of the given size with approximately equal frequency in the long run. The set of swans we happen to have observed, the parts of the universe we happen to have observed, the set of moments during which we looked at the paper in Stace's desk, the set of instants during which we see Chisholm's cat on the roof, are none of them selected by a method which is capable of replication, and which, in the long run, would select each possible equinumerous sample equally often.

But I would argue that this notion of randomness, though it is given a great deal of lip service, particularly by down to earth statisticians, is not only irrelevant to the consideration of inductive arguments of the sort here considered, but also to the most ordinary kinds of statistical inference, and even to the most common sorts of application of statistical knowledge. Randomization in this technical sense is neither necessary nor sufficient as a condition of valid statistical inference. If a sample is drawn from the hold of a ship, for example, for assessing the quality of a bulk cargo, the sample will typically and most usefully consist of smaller samples drawn from various parts of the hold. The sample is not one which is drawn by a method which would in the long run select every possible sample from the ship with equal frequency. The same example will illustrate the insufficiency of randomization: suppose the volume of the hold of the ship were divided into a very large number of very small volumes, that each of these very small volumes was assigned a number, and that a table of random numbers was used to select a large number of the very small volumes to constitute a sample of the cargo. Such a selection would in general perhaps be acceptable; but in a measurable number of cases, the selection would be rejected as an appropriate basis for inference on account of its bias. That is, sometimes this randomization procedure would result in selecting as a sample from

the hold a number of volumes along the keel of the ship, and none from any other part of the hold. Clearly such a sample, though as random as you please in the statistical sense, would not be random in the sense of being, relative to what we know, as good a candidate as any other sort of sample for being representative. Stratification is beside the point: no population and no stratum of a population is perfectly homogeneous, and thus randomized sampling in any population, or any stratum of a population, may result in a selection of a sample which is clearly less likely than other samples to be representative, relative to what we know about it. Randomness in the statistical sense is thus neither necessary nor sufficient for soundness of inference and is beside the point in the arguments at issue.

The question of statistical randomness versus epistemological randomness does not appear as an explicit problem in any of the discussions I have referred to earlier. In each of these discussions, the form of the argument has been: there is no inductive argument that leads from premisses of type *A* to a conclusion of type *B*. From the nonexistence of such an argument, interesting philosophical theses are alleged to follow. My refutation, in each case, consists in producing just such an inductive argument. It is clear from the tenor of the discussions I have been commenting on that the author would tend to take a certain tack in attempting to rebut my proffered inductive argument, and so I have gone one step further, and indicated how and why, in my opinion, that attempt would fail. From the existence of inductive arguments of the sort whose existence Broad, Stace, and Chisholm deny, certain philosophical theses also follow. These theses, e.g., that material objects exist, that perceptual judgments are usually veridical, that we have reason to suppose that metabolic reactions work the same way on the other side of the moon as here on earth are neither so startling nor so interesting as the contrary theses that these authors are led to. But it is my belief that they are more soundly based theses.

VI. OBJECTIONS TO PREDICATIVE RELATIONS

NICHOLAS WOLTERSTORFF

VARIOUS philosophers, from ancient to contemporary times, have argued that there is some incoherence in thinking of something as standing in the relation to a certain property of *having* or *possessing* or *exemplifying* it; and some incoherence in thinking of two or more things as standing in the relation to a certain relation of *standing in relation to* it. There is, they have said, no such relation for entities¹ to stand in to properties as that of *having* or *possessing* or *exemplifying*. And there is, they have said, no such relation for entities to stand in to relations as that of *standing in relation to*. I wish to look at a view of the more influential of these claims. Of course, any argument designed to establish that there are no relations at all will, if successful, establish that there are no such relations as those of exemplifying and of standing in relation to. It is not my intention here, however, to discuss such general arguments; but rather to discuss those which are designed specifically to show that there is no such relation as exemplifying and no such relation as standing in relation to.

Plato was the first to present an argument along these lines; and the problem he pointed to has worried many thinkers since. This is how Plato states his perplexity in *Philebus* 15b: "Whether, when one form is in many things, we think that the form is dispersed and has become many, or that it is entire and separated from itself—which latter would seem to be the most impossible notion of all—the problem remains: How can the same one form be at the same time in one and in many things." Examples which Plato here gives of forms are The Beautiful, The Good, Man, and Ox. The very same difficulty is raised by Plato in *Parmenides* 130E–131E. There the examples Plato gives of forms are beauty, justice, largeness, and similarity. He says that if various distinct things can share (*metalamano*) one of these forms, then we are confronted with the puzzle of how one form can be in many things. Cook Wilson, rather felicitously, calls Plato's

problem the problem of the unity of the universal in the plurality of its particulars.

It is perhaps disputable whether it is *always* satisfactory to substitute for Plato's phrase "the beautiful" (*to kalon*) the English "beauty"; for his phrase "the good" (*to agathon*) the English "goodness"; etc. But if we compare the *Philebus* passage with the *Parmenides* passage, it is clear that Plato thinks that the problem he is confronting here can be raised as well in terms of beauty as in terms of The Beautiful, as well in terms of goodness as in terms of The Good; etc. I shall, then, consider it in terms of properties.

We would, perhaps, not naturally speak of beauty, and goodness, and justice, etc., as *in* things. But for Plato, in *Parmenides* 130–131, it is clearly the case that if some thing partakes of (has, possesses, shares, shares in, has a share of) justice, then justice is *in* that thing and conversely. Plato, in fact, there states his puzzlement both in terms of how it can be that many distinct things each have a share of one form, and how it can be that one form is in each of many distinct things.

Plato's question is undeniably cryptic. But quite possibly one example of the general question over which he is puzzled is this: Suppose that *A* and *B* each possess paleness and are not identical; then does *A* have all of paleness or part of it, and similarly, does *B* have all of paleness or part of it (supposing all the while that there is only one such thing as paleness)? If this is indeed an example of Plato's general question, then Plato's comment about it is that he finds both answers to this question unattractive. The worse answer, he says, is that each has all of it; how can that be? But scarcely better, to his mind, is that each has only a part.

The partisan of universals need not be embarrassed by Plato's question. Indeed it cannot be that something has only part of paleness; for paleness is not the sort of thing that can have parts. But in allowing this, we are not thereby committed to

¹ Throughout, I shall need some common noun which is absolutely general in its application. I shall use "an entity" as such. Thus from any proposition of the form *x is a K* ("a K" here representing any common noun), the corresponding proposition of the form *x is an entity* follows.

holding that if something has paleness, then it has all of it. For it is surely consistent, and plausible besides, to hold that something has paleness without either having all of it or part of it. Paleness, it is plausible to think, is not the sort of thing that something can have all of, nor the sort of thing that something can have part of, while yet being the sort of thing that something can have. Alternatively, one could hold that if something has paleness then it has all of it, but repudiate Plato's assumption that two distinct things cannot each have all of it. Why should paleness not be the sort of thing that two distinct things can each have all of?

So far as we know, Plato's puzzlement over the concept of possessing never drove him to the conclusion that there can be no cases of one thing having or possessing another. Gilbert Ryle, then, in an article on Plato's *Parmenides*, goes a step beyond Plato; for he argues that this conclusion should be drawn. Ryle states his argument as follows:

Now what of the alleged relation itself, which we are calling 'exemplification'? Is this a Form or an instance of a Form? Take the two propositions 'this is square' and 'this is circular'. We have here two different cases of something exemplifying something else. We have two different instances of the relation of being-an-instance-of. What is the relation between them and that of which they are instances? It will have to be exemplification Number 2. The exemplification of *P* by *S* will be an instance of exemplification, and its being in that relation to exemplification will be an instance of a second-order exemplification, and that of a third, and so on *ad infinitum*. . . . This conclusion is impossible. So there is no such relation as being-an-instance-of. 'This is green' is not a relational proposition, and 'this is bigger than that' only mentions one relation, that of being-bigger-than.²

Before I try to state Ryle's argument here, let me explain by example how I shall henceforth use the word "case." I shall call *Socrates' wisdom* (alternatively expressed, *the wisdom of Socrates*) a case of wisdom; and so also I shall call *Socrates' being wise* a case of being wise. I shall call *Romeo's loving Juliet* (alternatively expressed, *the loving of Juliet by Romeo*) a case of somebody loving somebody; and so also I shall call *Romeo's love for Juliet* (alternatively expressed, *the love of Romeo for Juliet*) a case of love.

Now in the passage quoted, Ryle is apparently arguing thus: Suppose, he says, that there were such a relation as exemplification. Then

- (1) This is circular
would entail
- (2) This exemplifies circularity,
which in turn would entail
- (3) There is such a case as the exemplification of circularity by this, and such a relation as exemplification, and the former exemplifies the latter,
which in turn would entail
- (4) There is such a case as the exemplification of exemplification by the exemplification of circularity by this, and such a relation as exemplification, and the former exemplifies the latter,
which in turn would entail
etc., . . . *ad infinitum*.

Now, says Ryle, the relation of exemplification which the sentence used in (3) says that there is, and which "exemplifies" in the sentence used in (2) is thought to stand for, is first-order exemplification; that which the sentence used in (4) says that there is, and which "exemplifies" in the sentence used in (3) is thought to stand for, is second-order exemplification, etc., *ad infinitum*. In short, if there were such a relation as exemplification, then (1) would entail the existence of an infinitely large number of different orders (kinds) of exemplification, one of these being that which the sentence used in (3) says that there is, another, that which the sentence used in (4) says that there is, etc. But there cannot be such an infinitely large set of different orders of exemplification. So if "This is circular" did entail the existence of an infinitely large set of different orders of exemplification, it would have to be false. But it is true. So it does not entail this infinity. But if there were such a relation as exemplification, "This is circular" *would* entail this infinity. Hence, there is no such relation. (From which it follows, of course, that it is not a relation which, in making assertions, one claims to hold between things.)

It is not clear from our way, nor from Ryle's way, of stating the argument why he holds that the relation of exemplification which the sentence used in (3) says that there is, is different from that which the sentence used in (4) says that there is. What reason is there for holding that it is not the same relation which is said, with each of these sentences, to be, and thus for holding that (1) entails the

² G. Ryle, "Plato's 'Parmenides'," *Mind*, vol. 48 (1939), p. 138.

existence of an infinitely large set of different orders of exemplification?

Let us speculate a bit as to what Ryle might have had in mind here. Consider the following series of cases of exemplification:

- (iii) The exemplification of circularity by this
- (iv) The exemplification of exemplification by the exemplification of circularity by this
- etc.

Now compare that series of cases of exemplification with the following series of relations:

- (III) Exemplification of ____ by ____
- (IV) Exemplification of exemplification by the exemplification of ____ by ____
- etc.

The relation between these two series is obvious. (iii) is a case of (III), (iv) is a case of (IV), etc.

Now it will be noticed that case (iii) is what, in (3), is said to exist and to exemplify something; case (iv) is what, in (4), is said to exist and to exemplify something; etc. But since case (iii) is a case of relation (III), and case (iv) is a case of relation (IV), it immediately occurs to us that the sentences used in (3) and (4) are incomplete for Ryle's purposes. For (3) we should substitute this:

- (3') There is such a case as the exemplification of circularity by this, and such a relation as exemplification of ____ by ____, and the former exemplifies the latter.

And for (4) we should substitute this:

- (4') There is such a case as the exemplification of exemplification by the exemplification of circularity by this, and such a relation as exemplification of exemplification by the exemplification of ____ by ____, and the former exemplifies the latter.

Etc.

So suppose we state Ryle's argument in this way. Is it then clear that the relation which in (3') is said to exist is distinct from that which in (4') is said to exist? In other words, is it clear that the relation (III) is distinct from the relation (IV)? Granted that we now have a series of different *names* for the relation or relations said to be; is it clear that we have a series of different *relations*?

It seems to me that this is not at all clear. Certainly the relation (III) is equivalent with the relation (IV), in the sense that no two things could bear to each other the relation (III) without also bearing to each other the relation (IV), and vice

versa. Also it seems that to claim about the members of a pair that one of them bears the relation (III) to the other is to make the very same claim about them as that which one makes in claiming that one of them bears the relation (IV) to the other. Whether this is sufficient ground for concluding that they are identical relations is too large a question for us to consider here. Here let it simply be said that some argument for the non-identity of relation (III) and relation (IV), etc., seems necessary if we are to have a convincing proof that in the series of sentences used in (3'), (4'), etc., an infinite number of different relations is said to be. No such argument is offered by Ryle.

We said that (iii) was a case of relation (III), that (iv) was a case of relation (IV), etc. But now consider the following series of relations:

- (III') Exemplification of ____ by ____.
- (IV') Exemplification of ____ by the exemplification of ____ by ____.
- (V') Exemplification of ____ by the exemplification of ____ by the exemplification of ____ by ____.

Etc.

It seems clear that the series (iii), (iv), (v), etc., is not only a series of cases of the relations (III), (IV), (V), etc.; but is also a series of cases of the relations (III'), (IV'), (V'), etc. For example, (iv) is a case of (IV), but also of (IV').

Corresponding to this new series of relations, there is also the following new series of sentences:

- (3'') There is such a case as the exemplification of circularity by this, and such a relation as exemplification of ____ by ____, and the former exemplifies the latter.
- (4'') There is such a case as the exemplification of exemplification by the exemplification of circularity by this, and such a relation as the exemplification of ____ by the exemplification of ____ by ____, and the former exemplifies the latter.

Etc.

Now it does seem clear that in the series (III'), (IV'), (V'), etc., we have a series of *distinct* relations. For (III') is a two-termed relation, (IV') is a three-termed relation, (V') is a four-termed relation, etc. Thus, in the series of sentences (3''), (4''), etc., an infinite number of distinct relations is said to be.

But what, now, is impossible in this situation? Certainly it is not impossible that there should be

infinitely large classes, nor that a true proposition should entail the existence of an infinitely large class of things. All we have from Ryle's hand is the claim that this is impossible, without any clue as to why.

But even if one thought it impossible that there should be an infinite series of relations of the sort (III'), (IV'), (V'), etc., one would not have to conclude that there is no such relation as exemplification. For one could hold that while (2) entails (3), the succeeding purported entailments do not hold. In other words, one could hold that there is such a relation as exemplification, and that (1) entails (2), and that (2) entails (3); but that (3) does not entail (4).

In short, Ryle's argument does not offer us sufficient reason for concluding that there is no such relation as exemplification.

2. In a passage by P. F. Strawson, the claim is made, concerning *exemplifying*, and concerning *standing in the relation of something to*, that though there are such entities, they are not relations. This is what Strawson says:

Thus we use such forms as '... is an instance of ...', '... is characterized by ...', '... has the relation of ... to ...'. I shall appropriate some of these expressions, using them as the names of different kinds of asserted tie. ... It is important that we should not think of these two- or three-place expressions as themselves the names of terms of a certain kind, *viz.* relations. Something analogous to Bradley's argument against the reality of relations may be used, not indeed to show that relations are unreal, but to show that such assertible links between terms as these are not to be construed as ordinary relations. Let us speak of them as non-relational ties.³

Strawson here denies that "... has the relation of ... to ..." stands for a relation. He denies that there is any such relation as the relation of having the relation of something to. Yet he would allow that in assertively uttering "John has the relation of loving to Mary" we claim that John and loving and Mary are "linked" or "tied" in a certain manner.

Thus, as I understand him, Strawson would hold that from "John loves Mary" it follows that there is the relation of loving and that John has it to Mary; but he would deny that it follows further that there is the relation of having the relation of something

to. He wants, in other words, to distinguish between relations in general, and *having* or *bearing* or *standing in* the relation of something to. Standing in the relation of something to, he wants to say, is not a relation, or at least, not an "ordinary relation."

Now initially one might suspect that this is all merely a verbal matter. But in fact Strawson suggests, as his reason for saying that there is no such relation as that of standing in the relation of something to, that there is some absurdity involved in holding that this is a relation, or at least, in holding that this is an ordinary relation—an absurdity pointed out or suggested by Bradley. Strawson holds similar views for what the expressions "... is an instance of ..." and "... is characterized by ..." stand for. These do not, he holds, stand for genuine relations. They stand rather for non-relational ties.

Ryle and Strawson would agree that there is such a relation as loving. Ryle, however, would deny that there is any such thing as standing in the relation of something to, while Strawson would rather insist that though there is this, it is not a relation.⁴ Russell, then, goes both Ryle and Strawson one better, by denying that there are *any* relations. It may be true, he holds, that John loves Mary; but it does not follow that John stands in the relation of loving to Mary. There is some absurdity, he holds, in supposing that there is such a thing as the relation of *loving*. This is a view which he adopts in his paper on "Logical Atomism" as well as his book, *Philosophy*, and which in the latter he also attributes to Wittgenstein in the *Tractatus*. Russell says, for instance, "The first step in Bradley's regress does actually have to be taken as giving verbal expression to a relation, and the word for a relation does have to be related to the words for its terms."⁵ He also says, "The concept of the relation as a third term between the other two ... must be avoided with the utmost care."⁶ Russell appears to hold that one can never actually refer to any relation, and also that it is never true to say that there is such a relation as so-and-so. Russell's view says nothing against the possibility of one thing having or possessing another; but it does say that nothing ever stands in a certain *relation* to another thing. Things are related; but nothing stands in some

³ P. F. Strawson, *Individuals* (London, 1959), p. 167.

⁴ It is unclear whether Ryle thinks that "John loves Mary" entails "There is the relation of loving, and John stands in it to Mary." Quite clearly he *would* hold that there is the relation of loving.

⁵ B. Russell, *Philosophy* (New York, 1927), p. 253.

⁶ B. Russell, "Logical Atomism" in R. C. Marsh (ed.), *Logic and Knowledge* (London, 1956), p. 335.

relation to another thing. Russell's reason for denying that, for example, "John loves Mary" entails that John stands in the relation of loving to Mary, is his conviction that there is some absurdity involved in holding that there are relations, an absurdity pointed out by Bradley's regress.

To see, then, what it is that Strawson and Russell have in mind as lending justification to their views, we must look at the argument of Bradley to which both allude. To see the point that Bradley is making, we must quote him rather extensively:

One quality, *A*, is in relation with another quality, *B*. But what are we to understand here by *is*? We do not mean that 'in relation with *B*' *is A*, and yet we assert that *A* is 'in relation with *B*'. In the same way *C* is called 'before *D*', and *E* is spoken of as *being* 'to the right of *F*'. We say all this, but from the interpretation, then 'before *D*' *is C*, and 'to the right of *F*' *is E*, we recoil in horror. No, we should reply, the relation is not identical with the thing. It is only a sort of attribute which inheres or belongs. The word to use, when we are pressed, should not be *is*, but only *has*. But this reply comes to very little. The whole question is evidently as to the meaning of *has*; and, apart from metaphors not taken seriously, there appears really to be no answer. And we seem unable to clear ourselves from the old dilemma. If you predicate what is different, you ascribe to the subject what it is *not*; and if you predicate what is *not* different, you say nothing at all.

Driven forward, we must attempt to modify our statement. We must assert the relation now, not of one term, but of both. *A* and *B* are identical in such a point, and in such another point they differ; or, again, they are so situated in space or in time. And thus we avoid *is* and keep to *are*. But, seriously, that does not look like the explanation of a difficulty; it looks more like trifling with phrases. For, if you mean that *A* and *B*, taken each severally, even 'have' this relation, you are asserting what is false. But if you mean that *A* and *B* in such a relation are so related, you appear to mean nothing. For here, as before, if the predicate makes no difference, it is idle; but, if it makes the subject other than it is, it is false.

But let us attempt another exit from this bewildering circle. Let us abstain from making the relation an attribute of the related, and let us make it more or less independent. 'There is a relation *C*, in which *A* and *B* stand and it appears with both of them.' But here again we have made no progress. The relation *C* has been admitted different from *A* and *B*, and no longer is predicated of them. Something, however, seems to be said of this relation *C*, and said, again, of *A* and *B*. And

this something is not to be the ascription of one to the other. If so, it would appear to be another relation, *D*, in which *C*, on one side, and, on the other side, *A* and *B*, stand. But such a makeshift leads at once to the infinite process. The new relation *D* can be predicated in no way of *C*, or of *A* and *B*; and hence we must have recourse to a fresh relation, *E*, which comes between *D* and whatever we had before. But this must lead to another, *F*; and so on, indefinitely. Thus the problem is not solved by taking relations as independently real. For, if so, the qualities and their relation fall entirely apart, and then we have said nothing. Or we have to make a new relation between the old relation and the terms; which, when it is made, does not help us. It either demands a new relation, and so on without end, or it leaves us where we were, entangled in difficulties.⁷

As I understand him, what Bradley points out in the last part of this passage, the part containing the explication of the infinite process, is that, for example, in uttering the sentence "John is in the relation of loving to Mary" we cannot be understood as making the claim which is the conjunction of the claims that there are the entities John, Mary, and loving; nor the claim which is the conjunction of the claims that there are the entities John, Mary, loving, and the relation, call it *R*, which holds between John, Mary, and loving just in case the first is in the relation of the third to the second; nor the claim which is the conjunction of the claims that there are the entities John, Mary, loving, *R*, and the relation, call it *R'*, which holds between John, Mary, loving, and *R* just in case the first is in the relation of the third to the second; etc., *ad infinitum*.⁸

But, says Bradley, then he simply does not understand what could be claimed in uttering the sentence "John is in the relation of loving to Mary." For he has already considered, and discarded, the only other possibilities which occur to him. One possibility he considered is that "is in ____ to" is synonymous with "is identical with"; so that "John is in the relation of loving to Mary" can be paraphrased as "John is identical with the relation of loving Mary." But patently this is not the case. The other possibility he considered is that the person who utters "John is in the relation of loving to Mary" is claiming that there is some sort of tie-up between one or both of the terms and the purported relation. Perhaps he is claiming that one or both of

⁷ F. H. Bradley, *Appearance and Reality* (Oxford, 1930), pp. 17-18.

⁸ There are many other relations which Bradley might have in mind than the ones we have named *R* and *R'*. Instead of *R*, for example, he might have in mind the relation which holds between *any* two things and *any* relation just in case one of the things is in that relation to the other thing. But if Bradley's claim does not hold good for *R* and *R'*, it will not, I think, hold good for any of these other possibilities.

the terms *has* the relation (*has* the relation *to*), that the relation is an attribute of one or both of the terms. Perhaps he is *asserting* or *predicating* a relation of something, or *ascribing* a relation to something. In short, perhaps the sentence "John is in the relation of loving to Mary" can be paraphrased as "John *has* the relation of loving to Mary." Or perhaps as "John and Mary together have the relation of loving." But this, says Bradley, is no help. We need some explanation of "has"; but it is wholly unlikely that any satisfactory explanation will be forthcoming. For we are here confronted with the old dilemma concerning predication: "If you predicate what is different, you ascribe to the subject what it is *not*; and if you predicate what is *not* different, you say nothing at all." In short, Bradley seems to regard "John has the relation of loving to Mary" as not being a meaningful sequence of words, not a sentence which can be used to assert something.

A similar argument to this one can, of course, be given for any other sentence of the form "*x* is in the relation of *R* to *y*." And what Bradley concludes is that nothing true is asserted with such sentences. It may *appear* that something true is claimed in uttering them; but nothing is, *really*. And since Bradley quite clearly regards such a sentence as "John loves Mary" as no better than "John is in the relation of loving to Mary" he concludes that nothing is really related to anything else. "The conclusion to which I am brought is that a relational way of thought—any one that moves by the machinery of terms and relations—must give appearance, and not truth."⁹

The question to ask, now, is this: What is there about the infinite process which would lead Russell to hold that, though things are related, there are no relations, and Strawson to hold that, though one thing can stand in a certain relation to another thing, standing in the relation of something to, is not a relation? What Bradley says is that in uttering "John is in the relation of loving to Mary" we are not asserting any member of the infinite sequence of conjunctions indicated. How does this have any tendency to justify either Russell's contention or Strawson's?

Russell and Strawson do not wish to accept Bradley's *general* conclusion that nothing is really related to anything else. But the fact that in uttering a sentence of the form "*x* is in the relation of *R* to *y*" we do not assert any of the sorts of conjunctions

indicated, surely does not in any way justify the conclusion that nothing is related to anything else. Nor does Bradley think that it does. Rather, Bradley's argument is that nothing is asserted in uttering sentences of the form "*x* is in the relation of *R* to *y*" (or, of the form xRy). So far as he can see, nothing at all is claimed with them, certainly nothing true. It is from this that he concludes that nothing is related to anything else. Bradley's words strongly suggest the conclusion that his view is that the only claims he understands are identity and non-identity claims, and existence and non-existence claims. What he is apparently trying to do is to paraphrase affirmative relational sentences by using only affirmative and negative identity sentences, affirmative and negative existence sentences, and truth-functions of such. It is after failing in this attempt that he concludes that no claims are ever made in uttering affirmative relational sentences.

What can one say, to someone who holds that nothing is ever claimed in uttering affirmative relational sentences, but that one can see that sometimes something is claimed in uttering such sentences, and that sometimes what is claimed is true? What else can one say? Except that the person who argues that nothing is claimed in uttering affirmative relational sentences conducts his argument by uttering such sentences.

Now Russell and Strawson are clearly unimpressed with Bradley's general argument. Clearly, they believe that affirmative relational sentences are often used to make genuine claims. Further, they clearly believe that often, at least, they understand what those claims are; and that often those claims are true. But why, then, does Russell contend that though things are related, there are no relations? And why does Strawson contend that, though something can stand in the relation of something to something, yet standing in the relation of something to, is not a relation? These contentions just seem irrelevant to Bradley's actual argument. One can escape Bradley's conclusion without adopting either of them; and adopting either of them seems to provide us with no reply whatsoever to Bradley's actual argument.

Perhaps, though, Russell and Strawson have a somewhat different "infinite process" in mind, one suggested to them by Bradley's remarks, yet not quite the one which Bradley actually presents. Possibly they have the following sort of infinite

⁹ Bradley, *op. cit.*, p. 28.

sequence of *entailments* in mind (the letters *R* and *R'* standing for the same relations for which we had them stand above).

- (a) John loves Mary
entails
- (b) John is in the relation of loving to Mary
which in turn entails
- (c) John and Mary are in the relation of *R* to loving
which in turn entails
- (d) John and Mary and loving are in the relation
of *R'* to *R*
etc., *ad infinitum*.

In this case we have a sequence of entailments: (b) is the condition of (a), (c) of (b), (d) of (c), etc. If John is to love Mary, then he must be in the relation of loving to Mary; and if he is to be in the relation of loving to Mary, he and Mary must be in the relation of *R* to loving; and if he and Mary are to be in the relation of *R* to loving, he and Mary and loving must be in the relation of *R'* to *R*; etc. In general: For any relation (a relation being regarded as an entity), if some entities are to be in that relation to each other, those entities plus that relation must be in a certain relation to each other. It is this principle that generates what appears to be an infinite regress. Now perhaps it is the belief of Russell and Strawson that this principle is incompatible with the proposition that some things are related to each other. And since they hold that there are in fact things related to each other, they deny the principle. Russell stops the regress by denying that there are relations, and that one thing can be in a certain relation to another; he denies that (a) entails (b). Strawson stops it by denying that *R*, *R'*, etc., are relations; he denies that (b) entails (c).

Now this still leaves us with some puzzlement as to Strawson's total view. For though he holds that (b) does not entail (c), he apparently holds that (b) *does* entail "John and Mary are in the *tie* of *R* to loving." If so, do we not still have an infinite regress? And is not the distinction between ties and relations after all merely verbal? Possibly, though, Strawson's point is that there is no such entity as *R'*; and that "John and Mary are in *R* to loving" does not entail "John and Mary and loving are in *R'* to *R*." If this is indeed Strawson's claim, and if

furthermore it were true, then the regress would be stopped, and there would be a genuine distinction between what Strawson calls *ordinary relations* and what he calls *links*. The distinction might be put thus: Ordinary relations would be those relations which bear a relation to their terms. Links would be those relations which do not bear a relation to their terms.

But suppose that the principle enunciated does in fact generate an infinite sequence. Why should it be thought, on this account, that the principle is incompatible with the fact that there are things related to each other? If we may use an analogy—the principle that every natural number has a successor generates an infinite sequence; but it is not on that account at all incompatible with the fact that there are numbers.

Here again we can only speculate. But perhaps something like the following analogy is tacitly at work: Suppose that we wish to link *A* and *B* together. Suppose that *A* cannot be linked to *B* unless *A* is linked to the link, and *B* likewise. Suppose that *A* and *B* cannot be linked to the link, unless *A* is linked to the link which is to link it to the original link, and *B* likewise; etc. If all this were true, then it is clear that one is never going to succeed in linking *A* and *B*.¹⁰

Zeno already noticed that the movement from one place to another can also be made to look mysterious. Before one can go to *B*, one must go half the distance to *B*; but to do this, one must first go half of *that* distance; etc. But of course there is no incompatibility here. One can consistently hold both that space is infinitely divisible, and that we sometimes move. So too John can love Mary; even though, in so doing, he stands in the relation of loving to Mary, and he and Mary stand in the relation of *R* to loving, and he and Mary and loving stand in the relation of *R'* to *R*, etc., *ad infinitum*. In short, I see no incompatibility between the claim that things are related, and the principle that for every relation, if some entities are to be in that relation, then those entities plus that relation must be in a certain relation.

But I do not think that we have yet completely disposed of all that is perplexing about the relation of standing in the relation of something to. For clearly there can be the entities, John, loving, and Mary, even though it is not the case that John loves

¹⁰ *Ibid.*, pp. 27–28: "... the relation... , being something itself, if it does not itself bear a relation to the terms, in what intelligible way will it succeed in being anything to them? But here again we are hurried off into the eddy of a hopeless process, since we are forced to go on finding new relations without end. The links are united by a link, and this bond of union is a link which also has two ends; and these require such a fresh link to connect them with the old."

Mary. What then is missing? Loving was supposed to be a relation, something that related things. But there can be such a relation, as well as John and Mary, without the relation of loving doing its work of relating them. What is necessary, one wants to say, is not just that these three should exist, but that they should *stand in relation*. But there can also be this relation of standing in the relation of something to, along with John, Mary, and loving, even though it is not the case that John stands in the relation of loving to Mary. What then is missing? For again we seem to have a relation which fails to do its work of relating. It does, indeed, begin to seem futile even to speak of relations. What we need is ties which actually tie. So it seems. Or perhaps, better, what we need is just things in relation. And this is all we need—objects fitting into one another like the links of a chain.

The key to the dissolution of this perplexity seems to me to lie in the distinction between a *relation* and a *case* of that relation. We must distinguish between loving (the relation holding between any two things just in case one loves the other), on the one hand, and somebody's loving somebody (the loving of somebody by somebody), on the other. Loving is a relation. Somebody's loving somebody—e.g., John's loving Mary—is a *case* of that relation. We must distinguish between standing in the relation of something to (i.e., the relation holding among three things just in case one stands in the relation of another to the third), on the one hand, and something's standing in a certain relation to something (the standing in a certain relation to something by something) on the other. The former is a relation; the latter—for example, John's standing in the relation of loving to Mary—is a *case* of the relation.

We are apt to overlook the distinction between a relation and a case of that relation. We are apt to confuse the relation of loving, with somebody's loving somebody, to confuse the relation of sitting next to, with somebody's sitting next to somebody.

But especially we are apt, I think, to overlook the distinction between the standing in a certain relation by something to something, and the relation of standing in the relation of something to. Then only confusion can follow. For there can be such a relation as standing in the relation of something to, in addition to there being John, loving, and Mary, without its being the case that John stands in the relation of loving to Mary. Yet there cannot be such a thing as John's standing in the relation of loving to Mary, unless it is the case that John stands in the relation of loving to Mary. So if we do not see that standing in the relation of something to, is distinct from something's standing in a certain relation to something, then we are indeed perplexed. But the perplexity is eliminated if we keep in mind the distinction between a relation and its cases. There is such a case as John's standing in the relation of loving to Mary just in case John stands in the relation of loving to Mary. But it is not true that there are such things as John, Mary, loving, and standing in the relation of something to, just in case John stands in the relation of loving to Mary. The standing in a certain relation to something by something is not a relation, in particular, not the relation of standing in the relation of something to.

A case of a relation consists of things in relation; a relation does not. To get things in relation, we do not need some non-relational links or ties. All we need is *cases* of relations. If we just have relations and things, we do not have things in relation; but we do have things in relation if we have *cases* of relations. If all we have is the relation of love, and some persons, we do not have persons loving persons. For this, we need cases of loving.

We need not, then, deny that loving is a relation, nor that standing in the relation of something to, is a relation. Nor need we hold that relations are irrelevant to the phenomena of things in relation. For things in relation are just *cases of relations*.

VII. FORGIVENESS

JOSEPH BEATTY

THE intention of this paper is to provide an analysis of the forgiveness experience by employing as *paradeigma* Sartre's analysis of "Concrete Relations with Others" in *Being and Nothingness*. Thus, my paper hopes to indicate the fruitfulness of Sartre's method by extending it to the complex phenomenon of "forgiveness."

The experience of forgiving and asking for forgiveness is one of the most perennial of human experiences. Man's inability to live with others without offending, disturbing, and hurting them, and being subject to offense, disturbance, and harm at their hands posits the reality of conflict. It is not my concern, however, to explore the logical implications of conflict or to enumerate the presuppositions responsible for guilt following upon certain kinds of conflict and the concomitant appeal for forgiveness. Rather, it is my intention to examine the structure of experience usually designated "forgiveness." In examining this structure in as detailed a fashion as possible it is likely that the sensibilities of those who forgive and ask forgiveness will reveal judgments and presuppositions with regard to the conflict upon which forgiveness is founded. It is in exploring the actual experience of forgiveness, however, that such presuppositions will be brought to light.

Two sets of questions will guide my study of the experience of forgiveness: (1a) Under what conditions is it possible to ask for forgiveness? (1b) Under what conditions is it possible to forgive? (2a) What is one seeking in seeking forgiveness? (2b) What is one bestowing in forgiving?

Before proceeding to address the first set of questions one linguistic difficulty should be clarified. When we speak of "asking for forgiveness" and "granting pardon or forgiveness," it is easy enough to slip into the error of believing that forgiveness, because it is often expressed verbally or by means of a visible token of some kind, is something transmitted from one person to another. Forgiveness, however, is not negotiable as if it were a bond, not given over as if it were a commodity. It is not transitive in this sense. Indeed, the proverbial husband brings flowers to placate his

angry wife; such an action says "forgive me." The proverbial wife bakes the husband's favorite food to indicate that she has forgiven him. But such actions are emblematic of a change of attitude in the persons involved. The husband does not digest the wife's forgiveness with her food. On the contrary, it is the wife who *is* forgiving. Because we often articulate the experience of forgiving (and of offending) as if it were a structure composed of three terms—the person asking forgiveness, the person forgiving, the forgiveness—it is possible to think that forgiveness is transmitted from one person to another. In fact, we do often speak as though a person lacked forgiveness at one moment and was given it in the next. Because forgiveness is relational, one is often inclined to consider the forgiveness itself an element in the relation. This is a mode of thinking which Augustine employed in *De Trinitate* in his attempt to illuminate the notion of the Trinity by considering the love relation under the aspect of lover-beloved and the love between them. The same error is possible in describing the experiences of offending and forgiving and, in fact, we do speak of "giving offense" to someone just as we speak of "granting pardon" to someone.

The transitive or intransitive aspect of forgiveness is very important. For if forgiveness cannot be given, does this mean that it is futile to request it? On the other hand, what would it mean to possess the other person's forgiveness? Those whose love-relation has ended often find themselves with old love-letters which bear the beloved's written forgiveness, but her estrangement indicates her real lack of forgiveness. She has given her written forgiveness in her letter, but *she* was what the lover desired, not a piece of paper. Still, there seem to be both transitive and intransitive dimensions to forgiveness. The very appeal to the other to forgive one, which presupposes that one can influence the other to forgive, suggests that there is a very real transitive aspect to forgiveness. On the other hand, the suppliant does not desire something to be passed over to him (*viz.*, forgiveness); he desires a change in the other. It may well be that the appeal for forgiveness is transitive but the actual forgiving

intransitive. In this case, insofar as the circle of forgiveness is incomplete, the project of forgiveness is debilitated, and seems doomed to failure. On the one hand, the other cannot forgive unless she (or he) is asked to forgive; on the other hand, if she is asked to forgive she cannot truly forgive, for the request, in its active attempt to influence her forgiveness, may make it impossible for the forgiveness to be indeed *hers*. But *her forgiveness* is what the suppliant desires. If the person soliciting forgiveness in the very solicitation influences the other's forgiveness, he may well be (non-thetically) giving her the forgiveness with which to forgive him. In such a case the forgiveness (if it is finally conferred) is not hers but his own. What he desired, however, was not self-forgiveness but forgiveness by the other. But this is too short and sketchy a delineation of a paradox central to forgiveness. It will be completely considered as the questions proffered above are considered.

(1a-b) Under what conditions is it possible to ask for forgiveness or to forgive?

Prima facie, forgiveness would seem necessarily to arise out of a positive relationship of some sort. A consideration of forgiveness in a religious situation seems initially to corroborate this. The Scriptures tell us that the mercy and forgiveness of the Lord is infinite. Why is God merciful to man? Why is He always willing to "give him one more chance?" Because he loves or cares about him. Sin is conceived as an offense against God precisely because it is a violation of love, that is, an offense against a personal, providential Being, a loving Being.

Without this relation of love and care it is difficult to see how forgiveness could arise. If, like St. Francis of Assisi, one regards the birds and trees and flowers as brothers and sisters, one is careful how one treats them, and one is not indifferent to them, for they can be offended. Moreover, to ignore or maintain an attitude of indifference to one's loved ones is often an offense for which one must be forgiven just as if one had harmed them. For the religious man to ignore God is a sin for which he must ask forgiveness precisely because he has violated a bond of love. In short, one does not ask forgiveness of those about whom one does not care. Lovers' quarrels and the experience of "making-up" which involves forgiveness are possible only among lovers. Such forgiveness presupposes a relationship of concern which existed or exists between the two people.

In the light of this, the Christian commandment

to "forgive one's enemies" is puzzling. For to experience them as enemy one must be offended by them. But for them to offend us we must care about them to the extent that we can be offended by a less than positive action on their part. That is, for us to forgive an enemy, he must, in some sense at least, be or have been a friend, even a potential friend. We must have expected better treatment from him. For me to be offended by, and later forgive, a complete stranger who shoots me on the street, I must have expected better treatment at his hands.

But what of the man who spills coffee into my lap at a party and says "forgive me"? He didn't care about me one way or the other before his clumsiness. In fact, he may even have thought me a dolt and avoided conversation with me. But once, purely by accident, spilling the coffee upon me, he asks me to forgive him. Experiences of this type are many: the man in the crowd who steps on my foot, the woman who inadvertently bumps into me on the street, the painter who drops paint upon me from his scaffold. However, even in these situations aren't we often tempted to say: you should have been more *careful*? That is, don't we in fact make a tacit claim on the other's beneficence before such experiences? Don't we expect good treatment at his hands and remind him of it if he forgets to accord it to us? Who has not witnessed the spectacle on the streetcar when a man irretrievably drunk seats himself next to a respectable, self-righteous matron? Soldiers fighting on opposite sides do not need forgiveness at one another's hands for there is no relation of concern obtaining. In fact, there is a careful, premeditated carelessness about the lives of opponents. Only later, if they view their former killing in a different light, *e.g.*, as a violation of brotherhood, do they feel the weight of guilt and seek forgiveness. On the other hand, the betrayal of a soldier by one of his own men (he from whom the betrayed man expected benevolence) is felt to be a grave offense upon which guilt and the need for forgiveness follow.

The precondition for forgiveness, then, is the existence of a positive relation which is disturbed and often brought to awareness by the offense itself. Let me move now to the next two elements which constitute the situation out of which forgiveness can arise: an offending person and an offended person. These imply both guilt on the part of the offending person and the confirmation of the guilt by the person before whom the guilty person feels guilty. This is merely to say that if one of the two persons

does not believe that a genuine offense has been committed, the request for forgiveness is foolish or histrionic. If Falstaff were to say he felt guilty after stabbing Hotspur, already dead, we would think his confession of guilt mere histrionics. On the other hand, if the person before whom the guilty man feels guilty does not confirm his guilt by recognizing the man's offense as an offense to him, any forgiveness conferred would be bereft of meaning. In such a case, the man would forgive him to "humor him," as the saying goes, or in an attempt to dismiss quickly the thought of an offense. He would say, "There is no need to make a fuss since you didn't offend me in the least, but since you insist, of course I forgive you." But such a forgiveness is not authentic. What he really means to do is not to forgive the man (for he does not feel offended by the man), but rather, by a forgiving statement, induce the man to forget the offense he believes he has committed.

But this raises another very important question. If the experience of forgiveness requires as its condition both an offended and an offender, does this mean that the recognition of an offense arises simultaneously in both parties? It would seem that the conviction that one has been offended posits the responsibility and guilt of the offender. Similarly, the feeling of guilt on the part of one person in a relationship implies that the other has been offended, even if he or she is unaware of it. In such a case, the person asserts: I am guilty; you have been offended. Or, I am offended; you are guilty. But one significant refinement should be made here with regard to the guilt positing offendedness and the offendedness positing guilt. It is this: the claim that I am guilty really asserts that you *should feel* offended. And the claim that I am offended really asserts that you *should feel* guilty. This is an important recognition since it denies the necessity of guilt and offendedness arising simultaneously and it asserts the transitive nature of the guilt-offense experience. I shall return to this.

It is certainly possible to feel offended without the person whom one feels offended by recognizing his responsibility and guilt. Similarly, it is often the case that a person feels guilty without the person before whom he feels guilty feeling any offense at his hands. Some versions of the popular ballad have Sweet William unaware of the offense that pretty Barbara Allan feels; other versions have Barbara Allan oblivious to her offense to Sweet William. Often it is evident that storybook lovers who die of love for the women who have spurned them are

mistaken. The ladies did not spurn them at all. Their deaths are more the result of poor judgment (perhaps poor perspectives or poor eyesight) than of any offense. They feel offended and blame the ladies. For the ladies to sacrifice their own lives (die for love also), they must decide to see themselves as guilty in the offense which their men (whom they now decide to view as "lovers") felt and responded to by killing themselves. The lovers die for love with perhaps the intention of making the ladies *feel their guilt* in rejecting them and, therefore, guilty for their suicides. If the ladies succumb to this game and see themselves as guilty, then their deaths are an expression of their attempt to be forgiven by those they have decided to see as their spurned lovers. But Sweet William, in dying, implies that he *forgives* Barbara Allan. In fact, he makes a special request in the ballad that his friends not hold her responsible, "be kind" to her. He forgives her by affirming the judgment he thought she made concerning him (rejection) by taking himself *out of her sight* forever. He feels offended at her rejection of him but by confirming her judgment of him by removing himself from her sight, he acts as though she had judged rightly. He has made her rejection of him into a rejection of himself and thereby confirmed her alleged negative judgment of him. This action of his own transforms her wrong into a right. By his death he has translated her offense into a benevolent gesture by acting in accord with her offensive judgment and taking himself out of her sight. However, this very taking of himself out of her sight is, *ipso facto*, a placing of himself back into her sight. For his death itself confronts her with her guilt and her need for forgiveness. Her own death is an attempt to gain forgiveness but this attempt is futile since he at whose hands she seeks forgiveness is not available. It may be that she seeks self-justification and forgiveness in the eyes of the community. But even the elements judge her harshly and she becomes not a red rose, like the valorized Sweet William, but a briar. That is, she gains not forgiveness, but more accusation.

This is all by way of elaborating a case in which guilt and the offense did not necessarily arise simultaneously. In such a case, and in many other cases, "I am guilty" is equivalent to "You make me feel guilty," and "I am offended" is equivalent to "You make me feel offended." This is what I earlier referred to as the *transitive nature* of the guilt-offense experience. I can make someone guilty by being offended (making myself offended); I can make someone feel offended by feeling guilty (making

myself guilty). However, if the guilt and feeling of offense arise simultaneously might not this mean that the existence of a mutually recognized convention ("technique") is responsible for prompting the simultaneous guilt and offendedness? That is, have not both persons been made to feel guilty and offended by a convention which influences their reactions? For example, such commandments as "Thou shalt not kill" and "Thou shalt not commit adultery" would seem to give them their respective guilt and feeling of offense. They are not guilty and offended but have been made guilty and offended, not by each other, but by others. More properly, they have made themselves guilty and offended by responding to objective conventions.

With this discussion of guilt and offense in mind, let me now consider the central questions in my analysis of forgiveness:

(2a-b) What is one seeking in seeking forgiveness; what is one bestowing in forgiving?

In considering these questions I shall make use of some of the tentative suggestions and conclusions implicit and explicit in my treatment of the preconditions for forgiveness. It will also be necessary to consider the function temporality plays in structuring the possibilities and, *ipso facto*, the limits of forgiveness.

In delineating the situation in which one asks for and confers forgiveness, let me draw up an hypothetical case for the sake of concreteness. The following is a common enough case. A man offends the girl he loves by his inattention. He is involved in his work at the office and in cementing good relations with the higher-ups by taking them to lunch and playing golf, etc., with them. His girl feels neglected. This has happened before, she says, again and again. She has told him about it and he has promised to reform, but he never reforms and she sits at home. Other men are devoted to the girls they love, she says, but he almost seems to be consciously avoiding her or else—what is worse—he is forgetting her. Well, she has had enough of him and his neglect. She has undergone enough suffering at his hands. Finally she has "woken up" and no longer loves him, can never love him again, and therefore vows never again to see him. She tells him all of this in a letter.

In this letter we see all of the preconditions for forgiveness thus outlined: the relation of love obtaining, the offense felt by the girl, and her attempt to make her lover guilty. In making him feel guilty she attempts to convince him that it is

he who, by his neglect, has ended the relationship. She has merely formalized the breakup by telling him his offense. Thus, the scene is set for the man's appeal for forgiveness. For what does he ask in asking her to forgive him? But even before we can ask the question of the meaning of the man's appeal for forgiveness, we must examine how he feels when confronted by the girl's accusation and her rejection.

In this confrontation with his offense which initiates the situation of forgiveness, we immediately strike the paradox of the *offended offender*. Let us assume that he admits the reality of his neglect of the girl. Still, he feels that in the enumeration of grievances with which she confronts him she has offended him. For she has reduced him, by her accusation of neglect, to less than what he is. She has, in identifying him with only *some* of his actions, stood back and given him a characterization that is not comprehensive enough to include all of him. He has certainly been more to her than a neglectful beau. It is important to notice that the situation of forgiveness when it arises places the entire relationship in the past. The girl, in accusing and rejecting him and he, in his defense of himself, both refer to how they *have acted*. They objectify their relationship, and, by doing so, place it in the past. The man feels reduced to an object. He feels that he is fixed in place, motionless, by her accusation and rejection. In her conviction that he is guilty of these offenses to her, the girl has identified him with only some of his acts; in rejecting him because of his offense to her she tacitly posits his unchangeableness. Her offended pride makes him into a thing and rejects him. Her letter implies the judgment often accompanying the decision to terminate a relationship, "you'll never change." Both in her reductive identity of him with his neglect and her rejection of him on the basis of that identity, she confers upon him the status of an object in the past without the power to change. At the moment when she accuses and rejects him, he is pure past, pure facticity. Her vow "not to see him again" (i.e., not to look at him in the present) indicates this. She will not *look* at him; she will only *look back* upon him. The past is substituted for the future, memory is substituted for desire. Because she looks back upon him and judges him he feels reduced to an object just as she, in order to accuse and reject him, must feel that she has been reduced to the state of object by him: she feels reduced to one who has been neglected. She feels she is worthy of more, of fuller treatment, because she is more than his

neglect would make of her. It is her response to treatment as object which makes her an *offensive offended*. Her accusation is a refusal to admit the present. Indeed, this accusation is blind to his own sorrow and offense at the accusation. His remorse discountenances the accusation. Even if he admits the complete justice of her charge, if he is sorry for his acts of neglect he is already not identical with them. Thus, he is offended by her accusation of offense. He complains to himself and questions her value: if she loved me she would not stop loving me. All her promises to love him always are similarly falsified by her accusations. The future is collapsed into the past. The future he conceived with her has become impossible because he and she (the relationship) have become past. Also, he complains that if she really knew and loved him she would not reduce him to the past which he has already decided he is not, or rather, which he is, but *not only* that past. Furthermore, she is already different from the girl he knew and loved for that girl would never accuse and reject him. In a sense, the guilt and offense constitutive of the forgiveness-situation assert that both persons are *other than they were*. The forgiveness itself—the request and the conferral—represents an attempt to make things the same again, to seize the past as it was before the offense and accusation arose. But this is impossible because the girl in rejecting her lover on the basis of his past acts is also rejecting the past. She desires, as she might say, to begin again with someone else. Her lover tries to bring before her mind the past they had *before* and *besides* her offense, while at once owning up to and disowning his neglectful acts. He will attempt to make her “save” the past with him by showing her that it is full of loving actions, more loving actions than neglectful actions. For her to forgive him she must acknowledge that she was *wrong* in identifying him with some of his past actions. But this turns the need for forgiveness back upon her.

The girl's revelation of her offendedness makes the man both guilty and innocent. In affirming her accusation the man is willing to admit that such acts were wrong and that he is responsible but, he also insists, he cannot be completely concentrated into these acts in the past. For in the past he was more than the acts which are the basis of her accusation, and his present response (*viz.*, his sorrow for those acts) indicates how he stands apart from both his acts and his past. He feels offended by her statement of offense; it is unjust in that it wrongly gives a meaning to his past and refuses to

view him as over and against his past. He is offended by her reduction of him to his past; at the same time he already wonders how he can forgive her present accusation so that he can continue to affirm the past he spent with her and make it a basis for their future relationship. The girl has already *othered* herself by the accusation, just as she has othered him by concentrating him into some of his acts. He almost doesn't recognize her or himself as she sees him. By placing him in a position in which he must ask for forgiveness, she at the same time places herself in the position of needing forgiveness at his hands. In being offended and making him guilty she is, in turn, guilty in trying to make him guilty, for, in fact, he is not totally guilty. In accusing him she has attempted to see herself through his eyes, and in rejecting what she took to be his perspective, she rejected him and placed him in need of forgiveness. In responding to her accusations he attempts to see himself through her eyes, and in rejecting her perspective, he places her in need of forgiveness.

From her view, he has made himself, by his offenses, other than he whom she loved. Her offendedness which is the basis for her rejection of him is also that which prompts his desire to seek forgiveness. For her offendedness signifies that he is not the man he once was in her eyes. She says to herself—if he could offend me again and again how could he claim to love me; he, at the same time, thinks—if she could be offended at me and reject me, she must not have truly loved me. As I have already shown, the offense creating the forgiveness-situation is only possible in a relation of concern or love. But, while the offense is situated in the context of a love-relation, by its very nature it points the two people away from one another.

In some relations the offense is revealed by the offended to the offender in order that the offender will feel his guilt and ask forgiveness. In these cases the offended reveals the offense so that he may forgive; the offended may indeed desire to obviate the distance created by the offense. Mothers often tell their children their offenses so that the children, being ashamed, will ask forgiveness. Then, in other cases, the offended reveals the offense to the offender in the hope that he will deny or affirm it; the offended is not sure how the offender will respond. In the more extreme case that I have chosen, the offended tells the offense not to solicit the request for forgiveness, but to use the offense as a basis to terminate the relationship. In choosing to consider the more extreme case as a point of

departure it is hoped that the structure of forgiveness will be more fully illuminated.

Now I can come to the central question in the situation of forgiveness. For what is one asking in asking forgiveness? I am assuming that in asking forgiveness the person affirms that he indeed committed the acts of which he is accused. But he denies—and this is intimately bound to his appeal for forgiveness—that he is *in* them. For if he were *in* his acts then an affirmation of his guilt in the acts and his appeal for forgiveness would amount to a denial of himself. He is not sorry for *being* himself but for *having committed* certain acts of neglect. The very appeal for forgiveness is a practical demonstration that he transcends his acts and past and is not identical with them. In seeking the forgiveness of the other, the offender is asserting both that he is and is not the man who committed the offense. For if the offended sees no difference between the man who offended her and the one asking forgiveness, she has no basis upon which to forgive. If she recognizes no difference, then the forgiveness she confers merely amounts to an acceptance of the offender for what he is, viz., an offender. Then, “forgive me” means “accept me for what I am,” an offender. Her forgiveness would merely be a confirmation of her masochism, her desire for treatment as an object.

In asking forgiveness then, the offender places the offending acts before the offended person and, at the same time, asks her to recognize that he (the offender) transcends his acts. This transcendence is not manifest mainly in the difference between past and present or in the variety of acts (e.g., I have not only offended you in the past, I have also pleased you) but rather in the act of asking forgiveness itself. For to ask forgiveness of an offended is to invite the offended to see me as *forgivable*. That would involve her recognition of a divorce between what I am and have been from my offenses. The offender says, “I have not only neglected you. I have many times shown careful attention to you, etc.” Or, he says, “I did not mean to neglect you, I was not thinking.” Or, “Look at me, I am truly sorry. I have changed and would never neglect you again.”

But what indeed does the project of seeking forgiveness by making the offended person see me as forgivable involve? I invite her to see me not in the way she sees me in accusing and rejecting me, but rather to see me as I see myself, or as she herself once saw me, or as others who like me see me. The latter invitation explains some of the

function of the go-between, the friends I summon to talk to her about me. I send her letters, showing her as much of my attentive self in them as I can. I surround her with gifts, recall to her places we have visited together, people we have known, things I (the attentive, loving I) have said. In short, I surround her with an atmosphere of me which contradicts her image of me that is the basis of her rejection of me. I bring the past (exclusive of my neglectful acts) before her eyes to make her feel guilty about her charge. In seeking a basis upon which I can ask her to forgive me, I attempt (consciously or unconsciously) to make her feel her guilt in reducing me to some of my past actions. I want her to say to herself that “he’s not like I thought he was.” To banish her reduction of me I attempt to concentrate myself into a pose diametrically opposed to her classification of me. I want her to change her mind.

In this sense, the project of seeking forgiveness is a project to make the other seek forgiveness. For in seeking to make her aware that I am not to be identified with her accusing image of me, I am also trying to make her aware of the wrong she does me by her poor vision of me and the rejection based upon it. Her offendedness is an attempt (consciously or unconsciously) to make me guilty, but my attempt to gain her forgiveness is, in turn, an attempt to make her guilty so that she will seek my forgiveness. Thus, what I really desire in seeking her forgiveness is for her to forgive me. But what I awaken her to do is to seek my forgiveness for accusing me wrongly and inducing me to seek forgiveness. To seek forgiveness is, then, to invite the other to seek forgiveness.

She, in turn (if she decides to forgive me by admitting her wrong and seeking my forgiveness), invites me to see her as forgiving. But it is my campaign to exhibit my loving self to her which made her forgiving. My very project in seeking forgiveness posits her inability to forgive me on her own. I must, by making myself into a loving person before her eyes and thereby making myself appear forgivable, give her the forgiveness with which to forgive me. What I would like would be for her to see me herself and forgive me *on her own*. This is, in fact, why I feel offended when initially accused and rejected by her. I say: “She doesn’t know me at all. Why doesn’t she see me as I am?” But, this failing, I must be the Muse who puts the words in her mouth. I try to transfer my love of myself to her. I try to make her see and love me as I see and love myself, or rather, as I would desire myself seen and

loved. But my goal is short-range and not sought in all company. Tomorrow I shall perhaps attempt to endear myself to my employers by putting myself forward (projecting myself) as one who neglects everything (even the girl I am attempting to placate now by my attention) to attend to business.

But if the upshot of my seeking her forgiveness is to gain her desire for forgiveness I confront a self-contradiction. For in making myself appear forgivable in her eyes what I am doing is giving her my view of myself. In giving her my view of myself I am also giving her the forgiveness with which I would forgive myself. I am showing her how to forgive me as I would forgive myself. I am giving her the forgiveness with which to forgive me. But what I desired was *her* forgiveness. My very seeking for forgiveness (as an attempt to make her see me as forgivable) renders impossible her own conferral of forgiveness. I shall never quite forget that she did not forgive me on her own, that I had to "make her" forgive me.

Her attempt, provoked by my project to be forgiven (whose aim is consciously or unconsciously to elicit her desire for forgiveness), is to seek my forgiveness by showing herself to me as forgiving. But for her to be forgiving is for her to invite me to be forgiving. In forgiving me she is, *ipso facto*, asking me to forgive her for initially making me seek forgiveness. I am frustrated once more in my goal to gain her forgiveness. I gain only her search for forgiveness. But by putting herself forward as forgiving she is tacitly accusing my own project to penetrate her offendedness by my appeals for forgiveness. For such appeals imply that she is not forgiving of herself but must be coaxed to forgive.

Thus, for the offender to seek forgiveness is to make the other see him as forgivable, and thereby, to make the other seek the forgiveness of the offender. The forgiveness of the initially offended person takes the form of a seeking for forgiveness. This forgiveness is an invitation to the other to be forgiving. But the other cannot forgive; he only seeks for forgiveness. That is what forgiveness is: a seeking for forgiveness. Therein lies the central

paradox of forgiveness. In attempting to make the other see him as forgivable and thereby give her his forgiveness of himself, he actually succeeds in transmitting to her a recognition of her own fault in accusing and rejecting him. What he gives her, therefore, is not his forgiveness but his own seeking for forgiveness. He cannot give her *her* forgiveness, though that is what he's after. Nor can he give her, in the final analysis, his forgiveness of himself. He can only give her his seeking for forgiveness and she can only give it back to him.

Thus, it would seem that the project fails. The appeal for forgiveness prompts not forgiveness but the guilt which occasions desire for forgiveness. Forgiveness is transitive and intransitive, but the wrong qualities are transmitted. In attempting to make the other feel his guilt the offended makes him feel his innocence which is also the basis for the desire to be forgiven. Similarly, the offender, in attempting to make the offended forgive him (to give her his forgiveness) merely succeeds in making her feel her need to be forgiven by him.

However, this conclusion, in insisting upon the frustrating tension of claims and counter-claims, ignores the possibility of an equilibrium growing out of mutual recognitions and adjustments. Thus the oscillation can (and usually does) phase out, bearing the practical fruit of reconciliation. But even without final reconciliation, the project of seeking for forgiveness I have delineated may well be successful in the sense that it negates the other's claim. Forgiveness occurs, but such forgiveness is a surrogate for the expectations of those involved in what is customarily called "forgiving" and "asking for forgiveness." The experience in its complex structure redefines the initiating concept and adjusts the operations (though not necessarily the intentions) of the suppliant and the offended. This is not to say, however, that this reconstruction will be admitted and the experience consciously restructured in cognizance of the reversals and recognitions which the intricate task of forgiveness urges.¹

Northwestern University

Received April 23, 1969

¹ I am indebted to Josiah Thompson for criticisms of the original draft of this paper which were helpful in rewriting it.

VIII. ON THE LOGIC OF JUSTIFYING LEGAL PUNISHMENT

REX MARTIN

I

HOW do we justify punishment? Even if we ask what looks like a simple, clear question, we find, on reflection, that it is neither simple nor clear. The question is ambiguous; it can be read two ways: (a) "Why have punishment at all?" and (b) "How much, and of what sort, should there be in a particular case?" There may be other ambiguities in the question as well but, even if there were none at all, the question would not be clear. For before we talk about justifying punishment, we need to do some preliminary work on the concept of punishment: we need to clarify what we mean by "punishment" and to elaborate on the logic of justifying punishment, when the term is used in that way.

In this paper I intend to talk exclusively about punishment under law; the locus of my topic is political-legal as opposed to parental, educational, natural, or divine punishment. Let me begin by providing a conventional formula of definition for punishment in this restricted sense (i.e., as restricted to a political-legal institution or practice): *punishment is the carrying out in a legal way of a coercive penalty attached to a law as the "price to be paid" by the violator.*

But why this definition? The formula suggests that a "price to be paid" is automatically and justifiably attached to every violation of law. Do we intend to say, then, that a violator is punished simply because he broke a law, or that the violation of law is something about which it is reasonable to barter? Although this is an intelligible way to talk, and in some contexts might be helpful, it is singularly unilluminating here because what we want is a justification for saying "If you break the law then you will be punished." To say that the violator is punished because he broke the law or because he was willing to "pay the price" begs the

question whether (and why) there should have been a coercive sanction attached to the law in the first place.

The basic issue is to determine the ground of the legal entitlement to coerce as a penalty. Accordingly, we should revise our original formula of definition to square with that issue: *Punishment is the provision in a law for a coercive sanction to be applied, to those who break that law, as a penalty for violation.*¹

How does this revised "definition" of punishment change the picture? For one thing, it characterizes punishment, not as the carrying out of a penalty, but, rather, as the legal provision for a penalty. What one "justifies," if and when one does, is the having of any such provisions and the having of certain particular ones. If punishment cannot be "justified" at this level, at the level of legal provisions, then it can never be justified at the level of the actual carrying out of the provision. For a second thing, in the revised definition the conceptual status of the coercive sanction is that it is a *general policy* attached to a particular law; whereas, in the conventional definition, its status was that of a *particular coercive deed*. And, for a third, in the revised definition, punishment is conceived as a general policy of coercive sanctions applying to a *class of persons*, i.e., "those who break that law"; whereas, in the earlier definition, it was conceived as the carrying out of a particular penalty against a *particular individual person*, i.e., "the violator." Finally, the point of view in each definition is different: in the first, punishment is the "price to be paid" by the *violator*; in the second, punishment is a general-policy coercive sanction "to be applied" by some class(es) of *public officials*.

Once the concept of punishment is suitably stated, then we can turn to the issue of justifying punishment so conceived. What is it we are justifying? We are justifying the having of provisions in law for coercive sanctions to be applied as

¹ One might regard my "definition" of punishment as incomplete. My formula is intended to explicate one feature of the practice of legal punishment; it might be said that the formula defines the definitive feature of the practice. For completeness' sake, one could add to the formula the following: Punishment so conceived *includes* the carrying out of the provision.

penalties for the violation of laws. The issue is whether there ought to be, in a sense of "ought" to be specified, coercive sanctions in law. What would count as a justification? I want to show, in this paper, what a justification would look like; that punishment *could be* justified is all I want to argue.

II

To this end I would note that there are three standard normative theories of the justification of punishment: the Reformatory, the Retributivist, and the Utilitarian-Deterrent. There are three normative theories but they have in common the belief that punishment has, or can be shown to have, a "General Justifying Aim."² Each of them is conceived to be a theory for specifying that "Aim" (as, respectively, moral reformation, the righting of wrong, the deterrence of pain-producing actions). Is it useful to voice any sort of dissatisfaction with this whole approach?

Punishment under law is a political-legal concept; so is that of its justification. You might say: "This is a truism." But it is not, attended to, nonetheless. The objection which I want to raise is a conceptual one: it has to do with the way the three standard normative theories have mapped out the issue of justifying punishment, with the notion itself of a "General Justifying Aim" of punishment.

What I want to argue is that this way of conceiving the issue of justification is essentially wrong-headed: each theory is committed to a model in its talk about punishment (moral reformation, etc.) which is essentially pre-political, defectively political; whereas each should be, as providing a justification of punishment, an articulation of *political* concepts.

Let me put this point directly. Punishment has no *general* justifying aim. By this I mean that punishment is not some independent principle or institution that could exist, and be justified, anywhere and at any time in the world of human affairs; specifically, it is something that is dependent on the existence of governments and laws. It is a political institution which is essentially implicated, in *concept*, with other political institutions. In these respects it is unlike pain and deterrent counteractions and threats, which can be pre-political and are non-political. These latter things are politically ulterior; they are not *per se* political principles or practices.

But punishment under law is. Punishment is always located in a political system—some system of political institutions and principles. It is always located in some *particular* political system. A justification of punishment is always system-located, within some particular political system. Hence, for the remainder of this paper, I shall refer, not to its "General Justifying Aim," but to the "System-located Justifying Aim" (or, better, "Function") of punishment.

To justify punishment is to display its *rationale*. Punishment, which is system-located, can be justified only by showing its "necessity" within a particular system of *political* institutions and principles. And this involves showing its systematic relationships within the particular political system. (I am not talking about any particular state but about any particular *idea* of the state.) A justification of punishment is "internal," i.e., by reference to the internal relations of principles and institutions within a particular political system. A justification of punishment shows its place, and the "necessity" for it there, within a particular system.

But *moral* justification? What of that? Well, punishment under law is a *political* concept and I am not sure that it has any moral justification. I do not assume that it has one, or must have one. That is, if we accept a methodological principle of separability here, we have no reason to think that it needs one.

What I intend to suggest by introducing a principle of separability at this point is roughly this: we can never assume that politics and morals are conceptually the same, or overlapping, or in any relation of superordination/subordination. We can never assume that there is a complete coincidence of politics and morals. In fact, we should never assume that there is any overlap; although it is permissible to argue to this, or to a stronger conclusion. What we must assume is that they are *not* necessarily connected or mutually implicated. Political judgments do not necessarily presuppose or entail or covertly smuggle in moral judgments.

There may be connections here; I am certainly not denying that they should be related. What I want to say is that they are logically distinct. Politics is not essentially, or logically, a special case of morals; and political theory is not an applied form of ethics.

I am against taking the Necessary Connection of Politics and Morals for granted. I am arguing in

² The term "General Justifying Aim" is H. L. A. Hart's. See his "Prolegomenon to the Principles of Punishment" in P. Laslett and W. G. Runciman (eds.), *Philosophy, Politics, and Society*, series 2 (Oxford, 1962), pp. 158-182, esp. pp. 160-161 and 164-166.

this paper on the presumption that it has not been granted. This is, as I have indicated, a methodological presupposition.

As a political conception, punishment does require a political justification. But this is the only kind it *needs*. Now it may well have, or come to be supplied with, some sort of moral justification; this I do not deny. But it is difficult to see, if what is being justified is in fact punishment, what moral justification could add and what it would look like. Once the job of political justification is done, it is difficult to see any role for independent moral judgment.

Let me put this another way. I doubt that any ultimate justification of punishment is possible; it appears to be possible only if we make the mistake of conceiving punishment as an independent, non-political principle or institution, like infliction of pain. An ultimate justification, whatever that might be, is possible only indirectly; it would involve some "absolute" justification of the particular political system in which something conceived as punishment has its systematic location. Now such an ultimate justification might be possible, and if possible, might be a moral justification. Hence, if there is such a thing as an independent moral justification of punishment, it is probably at the level of justifying an entire system, and only indirectly a justification of punishment. But if punishment is directly justified only internally to a political system and if moral justification could be indirect and external only and applicable only to the whole system, then it would be less misleading to say that punishment *per se* can have no moral justification.

Beyond these two ways of justifying punishment—(a) directly, by displaying the necessity of systematically locating something conceived as punishment within a particular political system and (b) indirectly, by justifying the particular political system in which something conceived as punishment is located—I cannot conceive of what is being required of "a justification of punishment."

What I have tried to argue is that we need an essentially political conception of punishment, that we must have this—be able to think this way—*before* we can talk of a justification of punishment.

My argument contains certain features which have not been made wholly explicit; I think stating them will serve to remove something of the paradoxical character of my line of analysis. There are two "hidden premisses" which are worth noting. (1) The essence of penalty at law is coercion: some-

body (some personage at law) acquires the legal entitlement to coerce the violator, to coerce him in any of a number of ways (to die, to be imprisoned, to pay a fine, to suffer corporal punishment, to lose certain rights, to undergo psychiatric confinement, to be instructed in his moral responsibilities, etc.). Punishment is properly conceived as a species of physical *coercion or force* and not as a species of the *infliction of pain*. (2) Punishment is an historic, political institution. What punishment *is*—its functional description—will change from system to system.

My argument is intended to suggest that the concept of legal punishment requires a political description, which is provided in the categories of "legal provision for coercion" and "penalty for violation of law." And, more to the point, the *justification* of punishment requires the specification of the particular political institutions and ideals in the context of which the historically and systematically determinate practice of punishment is to be located, its description completed, and its justification carried through.

I do not want to be taken as saying that, in the nature of the case, it is logically impossible to bring all cases (or practices) of legal punishment under the scrutiny of some comprehensive normative ethical theory. Indeed, I mean to leave that question open. What I do intend to suggest is that legal punishment as described, either in a formula of definition or in a functional, system-located description, cannot be brought *directly* under some superordinate moral norm. To bring it under moral judgment, it must be internally related to other political institutions, which are themselves under their political description the locus of the moral judgment.

My argument is intended to bring out the point that the full description of any actual practice of punishment involves reference to other *political* institutions such that punishment can neither be ultimately justified nor discredited in abstraction from the system in which it is located. It is in this sense that the political justification of punishment has a logical priority.

III

What I think is called for in any justification of punishment as a legal institution is to locate the practice of punishment within a political system—for this is the only way in which I think it can be properly conceived as punishment and, possibly,

justified. To illustrate my argument with respect to justification, I will construct a schematic political system. I want to show in doing this only how legal punishment could be justified there; I will not actually attempt to "do" the justification in this paper.

In my illustration, I shall draw upon the political analysis formulated initially by Hegel and developed along one of its characteristic lines primarily by T. H. Green. The basic view here is of men associated in a system of rights. Such an association, insofar as its constituent rights are generated and systematically related by a government, is a political one; and the rights so generated are political and legal rights.

What is a legal right? Well, first, what is a right? A right, in the sense in which I am using the term (as in "system of rights"), is something that may be claimed, if three conditions are met. What is being claimed is a certain line of conduct, or of forbearance, and (a) it is claimed that this should be agreed upon for everyone. (b) This claim is "socially recognized" or mutually agreed upon. (c) The line of conduct is agreed upon on the basis that it is part of the "good" of each or instrumental to it (e.g., freedom of speech or "life and liberty"). In other words, the line of conduct is agreed upon because it is in the interest of each and all the members: each claims it for himself and recognizes it for all others on the basis that it is good for the association of all to be associated in what is good for each.

Now we know that under law there are many "rights" (legal rights). But not all legal rights are rights in the sense specified in the previous paragraph (*not all laws state something claimed for each and all, claims in the interest of each and all*). Those that are I shall call "civil rights." Civil rights are a special case of legal rights; civil rights, all of which are legal rights, are the rights with respect to which men are said to be associated in a "system of rights."

Government is the agency, the only agency,

peculiarly charged with the maintenance of the system of rights. I assume the need for "maintenance." The assumption is not far-fetched. It is conceivable that someone might "violate" a civil right, i.e., a right as legally defined. If this were to happen, then the system of rights would need to be restored—we don't want rights infringed.

Now, punishment would seem to come in at this point. For it is possible that some sort of provision for coercive sanction attached to law as a penalty for violating law might prevent such a violation (infringement) of rights from occurring in the first place or, if it has occurred, might prevent its recurring. In the simplest sense we can say then that "punishment" (i.e., the provision of a coercive sanction, as a penalty for violation, in law) exists to maintain and, if need be, to restore a system of rights. This is its "System-located Justifying Aim" or "Function."³

Is punishment, under this conception, "justified"? How could we justify *any* provision at all for coercive sanctions attached to law as a penalty for the violation of law? To do this we would have to demonstrate the "necessity" for such provisions within the particular system of political institutions and principles.

I have already mentioned two features of this demonstration: (A) Rights might possibly be infringed or interfered with. (B) Interference with rights is to be prevented. One might say that punishment is justified if it does, in fact, help in the preventing of interference with rights. I mean that punishment as a political institution is justified if it does, in fact, help in the maintaining—and in the restoring, where necessary—of a system of rights. In short, it is justified if and only if it is effective to that end. But this is a rather weak criterion.⁴

Having provisions for coercive sanctions attached to law, as penalties for the violation of these laws, is *prima facie* justified under a suggested *stronger* criterion if and only if two conditions are met. The meeting of these conditions is necessary and sufficient to establish that punishment is *prima facie*

³ The "System-located Justifying Aim" of punishment is the maintenance of a system of rights. This is *like* the aim of the Utilitarian-Deterrence Theory. But it is not identical with it: that theory specified the deterrence only of net-pain-producing actions, of such actions as such; whereas, in this theory, the aim is to deter actions violative of civil rights. Although punishment *deters* in both theories, its conception is vastly different in the two theories: each theory tells us something different as to what is to be deterred and why.

⁴ It is also ambiguous. The question really is: (a) Are we justifying punishment as a *political institution* by reference to its maintenance value for an *entire* system of rights? (b) Are we justifying a particular legal provision (for a coercive sanction, etc.) by reference to its maintenance value for a determinate civil right within the system of rights? I shall answer only question (b). But I think that the *kind* of answer offered to the question of maintaining determinate rights is of the right sort to provide an answer to the question of maintaining an entire system of rights. I shall assume, moreover, that the way to show that punishment helps maintain the *system* of rights is to show that it helps maintain various determinate rights.

justified in that *particular* political system (i.e., the one I called earlier "men associated in a system of rights"). (1) The provision of coercive sanctions in law will be effective in some degree to maintain the non-violation of a civil right within the system of rights. (2) No alternative to such provision will be effective or no alternative will be as effective.

Let us imagine that the basic *alternative* to provisions for coercive sanctions in law is some combination of the following: (a) public education, (b) reformist public policy, i.e., redistribution of social roles, economic goods, etc., and (c) public decision-making institutions, e.g., universal suffrage and regular elections. Let us assume that punishment is not included within the "combination" and that various combinations are possible under the headings (a), (b), and (c). Hence, there are a variety of alternatives to punishment for the *maintenance* of a civil right. We might put punishment and these alternatives on a list to be headed "General Social Control for the Maintenance of Rights."

Of course, it is also possible to include punishment within these various combinations. When I speak of an option as an alternative to punishment, I mean that the practice of punishment does not constitute any part of that option. The question whether punishment is justifiable is asked, over against these alternative options, with reference to its relative effectiveness, alone or in combination, in maintaining a variety of particular, determinate civil rights.

In other words, once we grant (A) that rights might possibly be infringed and (B) that interference with rights is to be prevented, punishment is *prima facie* justified as an institution within a political system—i.e., a system of rights—(C) if it would help maintain determinate rights within

such a system and (D) if it was *necessary* either in the sense that no alternative way would do that job or in the sense that no alternative could do the job as well.⁵

IV

I have provided a procedure under which the institution of punishment (i.e., provisions for coercive sanctions in law as a penalty for violation) could be said to be *prima facie* justified. Why do I say *prima facie* justified? Why not just justified, period?

Let us consider a rather common case, where a provision for penalty fails to deter and the carrying out of the attached penalty—in the form of some sort of coercion—is required. Should we in all such cases actually *use* force? Essentially, the answer is included in the one already given: if the *use* of force, as a penalty, is generally helpful in restoring a system of rights (i.e., helps deter subsequent violations of a determinate right) and if it was necessary then it would appear to be justified.⁶

But there would appear to be, I would suggest, certain infirming counter-conditions, against which the provision for using coercive force must be tested. And these counter-conditions are called into consideration against the *actual use* of force, the legal provision for which has been *prima facie* justified.

What I want to suggest is that the *prima facie* justification must be tested against various infirming counter-conditions, the failure to test against which infirms the final justification. The successful application of this apparatus—*prima facie* justification plus a test-out against infirming counter-conditions—to various cases of the practice of punishment would justify legal provisions for coercive sanctions. It would justify the practice in

⁵ The procedure for justifying punishment under this scheme is not circular. Even though the protection of rights might itself be claimed as a right, the practice of punishment is not thereby automatically a right. The practice of punishment is a claimable right only if it is justifiable and it is justifiable only if it is effective in limiting the violation of civil rights. By the same token, once the institution of punishment is established as politically justified within a system of rights, particular procedural safeguards, etc., can be claimed, with respect to penalty codes, as rights under law by the citizens. But again this involves no circularity as regards the justifiability of the practice of punishment. There is no right to protection by a penal code, and no determinate rights under this code, unless the practice of punishment is justified in the first place.

⁶ Again its "necessity" would require its effectiveness as a deterrent to subsequent violations and would be measured against alternatives: either no alternative could *restore* the system at all or any alternative was less effective than the use of force (under a provision for coercive sanction in law). Let me provide a brief list of alternatives, all of which are *by definition* non-coercive and all of which must be "triggered" by a violation of law to the end of deterring further violations of that civil right.

1. redress by public commissioners, etc.
2. blame
3. ostracism, shunning
4. voluntary education with goal of reform or improvement
5. voluntary restitution out of pure remorse (without blame by anyone else)
6. statement of general expectation of decent behavior: evokes apology, etc.

those cases without regard to the issue of the kind and degree of penalty, a question which can arise now and only at this point.

My point about calling these counter-conditions into consideration depends on the distinction between having provisions in law for coercive sanctions and the actual use of force in accordance with these provisions. In a sense, the distinction is artificial; it is a distinction that makes no difference. Most laws in point of fact have their effect, their maintenance value in the example at hand, through a combination of provisions for penalty and the employment of penalty; I have already indicated that, conceptually, any justification respecting the having of such a provision should *include* the case of its being carried out.

Nonetheless, there is a reason for marking the distinction. For if we were, simply to emphasize a point, to allow a distinction between merely having a provision in law and carrying it out, it would be clear that the latter raises at least two substantially new issues. First, in a system of rights, the use of force in particular cases would in all likelihood involve an "interference" with the civil rights of the violator himself (i.e., a right which he would normally have as a citizen is infringed upon or annulled by way of penalty). Secondly, the effectiveness of employing a particular mode of punishment is difficult to test and, at least in a system of rights, it seems that some sort of margin for error should be provided.⁷ There are fundamental reasons, then, why counter-conditions might be called into consideration against the actual use of force.

What are these counter-conditions? Let me provide some examples. (1) If the right violated is not thought to be particularly important or if its exercise is not particularly widely threatened, then it might be argued that the use of force is unnecessary or undesirable. (2) The *cost* of the *use* of force might be too high measured against the civil right violated, as regards the loss of rights on the part of the penalized violator. (3) The use of force might be or appear effective, but only to a limited degree. It has the apparent effect of widespread deterrence but it does *not* deter substantially (but only marginally) the particular person or class of persons actually and normally penalized. (An example from

outside the realm of civil rights would be laws against chronic alcoholism or against incest.) In any of these cases the provision for coercive sanction might be challenged.

It is important, in the analysis I have given, to separate the point about the role of these infirming counter-conditions in the logic of justifying punishment from the particular example, i.e., the case of a system of rights, which I have been using for purposes of illustration. The logical point arises, I have suggested, in drawing a distinction between the having of provisions for penalty and the use of force in the carrying out of these provisions. This distinction can always be drawn, regardless of the peculiar system-located aim which punishment serves; but it is not clear that the drawing of the distinction will always generate the sort of problem I have specified (where the penal use of force manifestly deprives the violator of some of the very benefits which it was the object of punishment to maintain in general and at large). My point, nonetheless, is that it is part of the logic of justifying punishment to specify the nature of the problem, to consider whether the problem does arise or not in a particular system, and to indicate the general character of its resolution in the notion of relevant counter-conditions and the test-out apparatus.

V

Let me summarize my analysis. I have tried to indicate when punishment (i.e., the provision of coercive sanctions as a penalty for law violation) is *prima facie* justified as a political institution; I have indicated by way of example that it would be when it is effective in maintaining determinate rights within a system of rights by deterring violators and when it is "necessary" as the relatively most effective device among alternatives. I have tried to indicate that any such *prima facie* justification can be infirmed in certain cases, when the actual use of force fails to test out against certain counter-conditions. But where the use of force as a penalty is not infirmed by counter-conditions, then the *prima facie* justification of punishment is ratified; punishment can be said to be justified in that particular political system.⁸

⁷ Such factors as the difficulty of testing *alternatives* to punishment, the presumption in favor of existing procedures, and the extrapolation of "proven effectiveness," such as it is, to the yet untested case are additional reasons why counter-conditions might be set up as infirming tests. Inasmuch as the initial determinations of effectiveness are themselves often sketchy and provisional, the counter-condition tests function as establishing a benefit-of-doubt zone.

⁸ My discussion throughout has been governed by the belief that it is reasonable to discuss the having of provisions in law for penalty sanctions without settling upon any precise rule for the kind and degree of penalty to be provided. My discussion presupposed that there was in each case some determinate penalty but I made no claim to the effect that it was the "right"

Now, although my conclusion—that punishment could be justified under certain procedures—is noteworthy, it is important to observe the scope and limitations of this conclusion with respect to the illustration I have been using. I have claimed that punishment can be conceived as a political institution within a political system, as an institution that deters the violation of rights within a system of rights, and that it might be justified in that role. This is the intended scope of my argument as regards that illustration. What limitations are to be observed here?

My argument was not designed to show that punishment might be justifiable in the case of *all* laws found in a system of rights. For some such laws do not support civil rights (i.e., some legal rights are not civil rights) and the sort of justificatory scheme offered—that punishment deters the violation of civil rights laws—would not serve to justify punishment in the case of those laws which are not civil rights laws. I have not, moreover, claimed that provisions for coercive sanctions might justifiably be attached to all civil rights laws as a penalty for violation. For I have argued, first, that there may be infirming counter-conditions directed against the

use of force to protect rights and, hence, that the provision for coercive sanctions, as penalties in those cases, could be effectively challenged and annulled. And, secondly, I would argue that it seems conceivable that a law could exist to state a norm (i.e., a claimable legal right) without provision for a penalty, e.g., the First Amendment to the Constitution. The mere fact that the law stated a civil right is not sufficient ground, although it is a necessary one, for saying that a provision for coercive sanction might justifiably be attached.

The question I set out to answer in this paper is “What would count as a justification of punishment?” I have, if only schematically, carried out my announced program: to show how punishment *could be* justified. I have not attempted actually to “do” the justification. That I have illustrated my analysis by referring it to a *system of rights* represents a philosophical and political prejudice, which I shall not undertake to defend at this time. In any case, I have provided the general outlines of a justification of punishment, indicating the sorts of things one would do if he were to carry through such a justification in any political system.⁹

University of Kansas

Received July 11, 1969

one. Such a claim would require that we have available a procedure, or set of rules, for such determinations. I would suggest that any specific justifying aim for punishment carries with it certain general guidelines appropriate to the question of deciding on kind and degree of penalty. Let me state briefly what I think these are in the particular case at hand (i.e., the case in which punishment has as its system-located justifying aim the maintenance of determinate civil rights): (1) Whatever is the kind and degree of penalty provided for, it should be consistent with the general maintenance of a system of rights (e.g., it should be consistent with the principle of equality before the law); (2) the penalty provided for and its employment should be effective toward maintaining non-violation of the determinate civil right and should not contribute to disrespect for rights-supportive laws or for the general system of rights; (3) penalties can be considered in combination with any other political devices, or set of such devices. Other criteria could be added (such as efficiency and welfare) but these are extrinsic to the notion of a system of rights and would not be entailed as guidelines by that notion.

⁹ A shorter version of this paper was read at the Western Division meeting of the American Philosophical Association in Cleveland, May 1, 1969. For the time and generous advice he gave me at the inception of this paper, I want especially to thank Professor N. H. Mucklow.

IX. EGOISM AND THE CONFIRMATION OF METAMORAL THEORIES

JOSEPH MARGOLIS

EGOISM is a much-maligned and neglected doctrine respecting the justification of one's conduct. By various strategies it is alleged to fall outside the pale of ethically relevant theories (though what the defining conditions of admissible theories might be is often unmentioned or, if mentioned, indecisive or prejudicial);¹ it is also sometimes thought to be inherently self-defeating or self-contradictory since the rational egoist cannot promote his doctrine among other men (though why he must or ought to do so or why the defensibility of egoism needs to be taken up only by egoists is ignored).² The point of exploring the tenability of egoism, apart from its intrinsic interest, lies in the quite instructive light it casts on the proposal of any putatively supreme justificatory principle of conduct. This alone makes it rather surprising that egoism has not been more attractive as a topic of controversy than the literature indicates, for it is by no means without charm or force or subtlety.

Traditionally, the particular bite of egoism has been assigned to its possible parity with utilitarianism: the argument, more or less as Henry Sidgwick has noted, lies with there being, for an hedonic criterion of value or for any comparable criterion (that might serve utilitarianism), an egoistic twin of any universalistic thesis. If this be admitted, the question immediately arises as to the basis on which to prefer the universalistic version (utilitarianism) to the egoistic version. It is obvious that, if egoism is either inadmissible on some technicality or incoherent or self-defeating, utilitarians are freed from the responsibility of demonstrating the superiority of their principle over some egoistic twin. Another alternative, the one adopted by Sidgwick himself and, generally speaking, by *laissez-faire* theorists is that egoism and utilitarianism happily come to the same thing. In our own

time, this is bound to appear merely naive: consequently, short of ruling egoism out for one reason or another, the question posed remains.

Now, an egoistic *element* can be shown plausibly to be an element of any tenable utilitarianism just as a deontological *element* can be shown to be an element of any tenable utilitarianism. Whether this means that utilitarianism is hopelessly inadequate as a candidate for the supreme justificatory principle of conduct pretty well depends on what is admitted to be the sense of the term. If, by strict utilitarianism, one means the thesis (in any of its versions) that what is morally right is what promotes the greatest good for all (where "greatest" and "all" are merely calculative terms and "good" designates some quality that lends itself to relevant calculation), then indeed if an egoistic element be acknowledged, the utilitarian principle will thereby be admitted to be inherently inadequate. If, on this reading, utilitarian calculation may, in principle, admit that egoistic preferences do, at times, outrank utilitarian ends, then of course one cannot consistently subscribe to the utilitarian principle as the supreme principle of morality. Consider, in this connection, that if the good produced by one's act be considerable and equal, addressed to oneself or another, and if the loss of value in not acting is similarly considerable and equal, and if the act cannot be performed for both oneself and another, it is not obviously unreasonable that one should prefer to perform the act in question for oneself rather than for another: on utilitarian grounds, on the hypothesis, there is no basis for preferring the one over the other—the matter is ethically indifferent; but on egoistic grounds, the matter is readily settled. Consider also that if the good produced by one's act be considerable and unequal when addressed to oneself or another, and if the

¹ I should make a particular exception of J. A. Brunton's instructive paper, "Egoism and Morality," *The Philosophical Quarterly*, vol. 6 (1956), pp. 289–303, in which it is explicitly noted that "what are generally regarded as minimum requirements for a moral system," namely, "over-ridingness, comprehensiveness, and the acceptance of rules of behavior" are, it may reasonably be argued, fulfilled by egoism.

² Cf. for instance B. Medlin, "Ultimate Principles and Ethical Egoism," *Australasian Journal of Philosophy*, vol. 35 (1957), pp. 111–118.

loss of value in not acting be considerable and unequal, and if the good that would be produced for another is slightly greater than the good that would be produced for oneself and if the loss of value suffered by another would be slightly greater than the loss of value that would be suffered by oneself, and if the act cannot be performed for both oneself and another, it would still not be unreasonable to act in one's own interest rather than to maximize goodness. Of course, a strict utilitarian of the stripe already defined will deny this, but the matter is arguable and it is not clear that it can be settled by an appeal, except viciously, to what may be supposed to be best under the circumstances. But if these counter-instances stand, they entail either that some ethically relevant choices are decidable only on non-utilitarian grounds without positively violating the utilitarian principle or that some ethically relevant choices are decidable only on non-utilitarian grounds with respect to which, precisely, utilitarian considerations are either violated or superseded. In either case, utilitarianism will prove to be inadequate as the supreme principle of conduct. All the more reason, then, for considering whether egoism might not be a strong candidate for such a supreme principle.

It can also be shown quite simply that a parallel problem for utilitarianism arises *vis-à-vis* a deontological element. Consider that one's acting in alternative ways produces two considerable and equal lots of good and that either alternative produces considerable and equal lots of losses of value, that the alternatives are exclusive, and that the distribution of good and loss is relatively equal on one alternative and is substantially unequal on the other. It may, not unreasonably, be argued that the alternative involving equal distribution is *prima facie* morally preferable to the other. But then, once again, a matter that would be indifferent on utilitarian grounds (as interpreted) would not be indifferent on deontological grounds and we should once again find it difficult to decide, except viciously, which alternative was morally more tenable. Alternatively, a comparable choice may arise where the utility of one alternative is slightly greater than that of the other although at the cost of considerable inequalities in the distribution of good and the loss of good. Under such circumstances, again, it seems not unreasonable to argue for the equitable distribution in the face of an over-all loss of utility. But then, of course, we should

be faced with deontological counterparts of the dilemmas posed by egoism. Short of a dogmatic rejection of such cases, we should find ourselves stalemated respecting the supreme principle of conduct. Also, it should be noted, some utilitarians would be willing to absorb either egoistic or deontological elements within their doctrine, though to the extent to which they would do so, they risk obscuring the lines of battle between alternative candidates for the supreme principle. Thus, for instance, one recent utilitarian declares that "considerations of equal distribution are central to the whole conception of utilitarianism."³ But this is simply to ignore possible discrepancies between maximal and equitably distributed utility; or, if it is willingly countenanced, the definition of utilitarianism will require conditions imposed on the maximizing of strict utility—which seems anomalous. In any case, partisans of the alternative views will have to be met and no conceptual gain will be affected by a mere change of label.⁴ The foregoing arguments, of course, presuppose the calculability in some sense of quantities of utility; but this is merely to concede to the utilitarian what is minimally required for his claim to be at all eligible.

Deeper objections to utilitarianism are not difficult to formulate. If the utilitarian objective is to guide rational calculation at all, it must be demonstrable that "the greatest good for all" or "the greatest good for the greatest number" is actually open to appropriate calculation. If essentialism or at least the thesis that man has a normative nature which may be assigned to him on cognitive grounds (implicit at the very least in Mill, for instance) is repudiated, it will be seen that the prospect of calculability rests on a number of extremely doubtful or flatly untenable assumptions. For, as soon as that thesis is denied (particularly if we keep the egoistic and deontological qualifications just considered in mind), it becomes entirely arbitrary to distinguish between the real and apparent goods of "all" (or "of the greatest number") or to sort out which "goods" sponsored by competing proposals are the ones to be maximized. But this is not all. For, on any preference (regardless of its content), it will have to be supposed by the utilitarian that the goods of all those to be considered are relatively *compatible*, so that a program of comprehensive planning is even remotely feasible; and it will also have to be

³ Cf. Jan Narveson, *Morality and Utility* (Baltimore, 1967), p. 229.

⁴ I find related and quite similar arguments in Nicholas Rescher, *Distributive Justice* (Indianapolis, 1966), ch. 2.

supposed by the utilitarian that the goods of all those to be considered are relatively similar and predictable, so that a program of comprehensive planning may be *relevant* and manageable; and it will further have to be supposed by the utilitarian that the ranking of all such goods exhibits at least the property of transitivity, so that a program of comprehensive planning may be rationalized and *calculable*. There is, however, no reason to think that *if* the values to be realized are to be moral values, that is, overriding values and not merely the presumptive values of prudent agents, that any of these conditions can be met *without the assumption of a normative human nature proper to man*. In the light of the historical record, there can be no doubt that the relevant (overriding) values depend on the appreciative preferences and tastes of different men in different ages. More sympathetically put, utilitarianism may be supposed to adopt, as its objective, the fulfillment of the presumptive interests of *prudential* agents (preservation of life, reduction of pain, etc.), which it then construes indefensibly as the *moral* objectives of man. But of course a rational agent may even suicide. The utilitarian, therefore, either on implied cognitivist grounds or on an implicitly normative reading of what it is to be "rational," inevitably confuses the objectives of certain conditional and technical interests with those that concern the overriding values of human existence.

To return to the main thread of the argument, let us grant, then, that there are grounds for supposing that strict utilitarianism must be modified to include not only certain deontological elements but also certain egoistic elements. The egoist may in his turn, however, claim that his principle does not stand in need of any such modification in favor of an otherwise competing principle. If so, egoism may very well exhibit a certain elegance and simplicity, and we may well wonder why it is not inherently preferable to utilitarian or deontological alternatives (for deontological theories are thought to be defective in their own right). Let us consider the possibility.

The egoist holds the view that the sole justification for his actions is that they contribute to his own interests, desires, pleasures, or the like. Precision is not required here as to whether the egoist is also a hedonist or prefers a conative criterion or the like: the issue is a perfectly general one and holds, *pari passu*, for all criteria otherwise eligible to its universalistic twin. There is, however, an interesting equivocation that arises here. One may com-

plain that the egoist is not advancing a universal principle, in the sense that he is not holding that everyone *ought* to subscribe to egoism and ought to be equally guided by egoistic considerations (or that, acting thus, men are acting "rationally" or "rightly" or in accord with what is "good"). If this is true for the egoist, a counterpart complaint may be laid against the utilitarian, for if utilitarianism is defective, then either a utilitarian is content to provide us with the principle on which he justifies his own conduct (in which case, he is doing no more than the egoist) or else he is merely advocating that everyone ought to subscribe to the utilitarian principle (in which case, he has yet to provide a suitable justification). In a word, unless one can demonstrate that principles like the utilitarian or egoistic or deontological are somehow true, on independent grounds (in which case, one may properly and neutrally claim that one ought to subscribe to the true principle of morality), we are confronted merely with the partisans of this or that principle and asked to consider their relative merits without regard to the question whether any one of them is straightforwardly true at all. It then becomes entirely irrelevant—apart from considerations of consistency—whether one advocates his preferred principle to the whole world or not. But this is an embarrassment, given the heat with which the relevant controversies have been waged.

On the other hand, the *objective* of the egoist is, indeed, his exclusive well-being whereas that of the utilitarian is the well-being of all—in this sense, then, a universalized objective. On this interpretation, the charge is true enough but seemingly question-begging, since this is precisely to rule out, at the very outset, the egoistic alternative and apparently to vindicate utilitarianism without a contest. The objection *may* stand, but it is obviously in need of supporting arguments; that is, the egoist's being occupied only with his own well-being may violate other (non-question-begging) conditions that defensible ethical principles may be expected to meet. The issue remains, what are these conditions? It may be emphasized, also, that the egoistic principle *is* actually universalizable in the further sense that everyone could, logically, subscribe to it just as easily as to the utilitarian principle. Also, the egoist is able to hold that, *if* his principle is true, then everyone ought to subscribe to it (if they are rational)—though in saying this, he need not advocate it and may, as a rational agent, even hope that others will not subscribe to it. Consequently, short of establishing that one or the other candidate

principle is true with respect to the moral domain, the only objective to egoism appears (thus far) to be that the egoist is prepared to justify his conduct solely on the grounds of advancing and maximizing his own well-being—that is to say, he is criticized merely for holding the principle he holds.

Now, the ultimate justification the egoist is prepared to offer is that he is uniquely himself, that what suits *him* justifies his conduct since no one else is, *ex hypothesi*, sufficiently like him (being different from him) to extend the justification to such another in suitably similar circumstances. Hence, the egoist may well claim to do as he pleases and on rational grounds; if, *per impossibile*, another were sufficiently like him, he would be obliged, on grounds of consistency, to regard as justified the other's behavior in like circumstances. In that case, he would of course fail to be an egoist, since he should then have failed to *justify* his exclusive concern with himself, however exclusive his actual concern may be. This suggests that we look more closely at the egoist's claim of uniqueness.

If one could show that all referring expressions could be eliminated and that such paraphrastic programs as Russell proposes in his *Theory of Definite Descriptions* or as Quine proposes in his attempted elimination of proper names were feasible,⁵ then, interestingly enough, the egoist would be easily defeated. For, if individuals were uniquely singled out by some set of indefinite descriptions and the like, the egoist would mean by saying, "Because I'm me" that he justifies his conduct on the grounds of possessing this or that attribute. But then, for any given attribute, accepting the doctrine of the "divided reference" of predicates,⁶ the egoist would be bound to *entertain* a non-exclusive justification of his conduct, since another might, logically, possess just that attribute (even if it is an exclusive attribute like that of being the tallest man in the world); also, should he happen to possess some attribute uniquely, it would still be necessary for him to show that the attribute singled out was *relevant* in justifying conduct in any sense at all. For instance, if our egoist were indeed the only man with nine toes on his right foot, it would normally be beside the point to mention this in defense of any particular action and it would certainly be a bore to have it offered as a justification for every action he performs. On the other hand, programs such as Russell and Quine have

proposed are by no means obviously feasible: the egoist is not, therefore, bound to subscribe to them and may well wish to justify his conduct in terms of *his* career, *his* interests, *his* pleasures. On that basis, he could, trivially, universalize his maxims, secure in the knowledge that they apply uniquely to him. Also, as we have seen, utilitarianism itself appears to require an egoistic element at least; but if so, some provision must be made for just the sort of consideration the egoist wishes to elevate to a supreme principle of conduct. Also, the egoist would never be obliged to face a challenge to the exclusiveness of his reasons for acting this way or that and, if there were *any* relevant reasons he might advance, they would all be selected from the set of reasons exclusively available to him. The only criticism that can be leveled against him, it seems, is that his reasons, curiously, will all be degenerate—that is, logically degenerate—since they will never really depend on this or that attribute but only on *his* possessing this or that attribute. As far as I can see, the egoist may be as fickle and changeable as he likes and still be said to act on rational grounds: he does what he pleases, but there may be no other significant regularities that may be singled out. We are, then, pretty well reduced to being spectators of the egoist's life, since as long as he is consistent there are no disputes possible with him, *on his grounds*.

Here, it seems, a clue suggests itself regarding the limitations of egoism. For, given the egoistic principle, there is no use (except, perhaps, a redundant or rhetorical one) for such terms as "right," "wrong," "ought," "obligatory," "forbidden," "duty," "rights," "just," and the like. The egoist is solely occupied with what, on his thesis, is good or bad. An interesting parallel may be mentioned here: Kant remarks that beings with "holy wills" cannot properly be said to be bound by duties, since such wills never deviate from willing what reason takes to be morally right; only imperfect beings, whose inclinations may go contrary to what reason tells them is morally right, can properly be said to have duties. By contrast, an egoist cannot consistently admit that he has duties because he cannot consistently admit any justificatory considerations that do not spring from his own interests (and, since rules are, for him, predictive only, rule-egoism must reduce to act-egoism). Consequently, an egoist simply does not share any of

⁵ Cf. P. F. Strawson, "On Referring," *Mind*, vol. 59 (1950), pp. 320-335; and Joseph Margolis, "On Names: Sense and Reference," *American Philosophical Quarterly*, vol. 5 (1968), pp. 206-211.

⁶ Cf. W. V. Quine, *Word and Object* (Cambridge, 1960).

the usual issues debatable by utilitarians or deontologists. For, not only is his objective his exclusive well-being but his reasons also are exclusive. It is, however, logically impossible to admit rights, duties, and the like without admitting that the justifying reasons for actions relative to these are, if valid, valid for like persons in like circumstances; that is, the very admission of questions of what is right or obligatory excludes the egoist from the debate, just as the adoption of egoism precludes such questions from arising. This, I believe, lies at the heart of the charge that egoism is not an ethically eligible position.⁷ The trouble is, once again, that the consistent egoist will merely maintain that, on his view, there are no independent matters of obligations, rights, etc. We are, therefore, driven back to our stalemate. Alternatively put, the limitations of egoism are not, it would appear, either moral or logical: rather they are practical, in the sense that the disputes of moral partisans predictably concern precisely those issues that egoism rules out of court. And this means that it is not the tenability, but the relevance, of the doctrine *vis-à-vis* the dominant issues, that will come under fire. At best, however, this is not even remotely decisive; and at worst, it is itself contingent on the stability of all those social influences that make of egoism a marginal theory. But the core difficulty of egoism must still be admitted, that is, that it makes no provision whatsoever for the debate of *public moral policies* as such, for it does not consider any recognizable attribute of any policy as contributing to a favorable appraisal (except in terms of an exclusive, personal interest).

Let me recapitulate, then, the main lines of the argument. The large alternative theories I am speaking about—egoism, utilitarianism, deontology—I shall call “metamoral” theories. I distinguish them from so-called “metaethical” theories, which are putatively restricted to the analysis of the key terms and locutions of ethical discourse. Metaethical theories are frequently assumed to provide an altogether neutral analysis of key terms and the locutions of ethical discourse employed in the substantive disputes they subtend. But it is not unlikely that one’s metaethical convictions (as that “right” is definable in terms of “good” or that “right” and “good” designate independent ethical concepts) are quite naturally affected by one’s substantive ethical convictions—which, therefore, correspond to them and differ significantly from the metaethical views of the partisans of alternative

ethical convictions. This is not to say that there are no relevant differences between metaethical and metamoral views: metaethical theories are theories about ethical discourse; metamoral theories are theories about the supreme principle by which ethical questions are resolved and conduct ethically judged and guided. It is to say, however, that one cannot identify *in an ethically neutral way* the primary data of the moral domain that *either* metaethical or metamoral theories are designed in their different ways to explicate. The view I wish to put forward here, in fact, is just that metamoral theories are not and cannot be *descriptive*, in any relevant sense, of the systematic justificatory features of moral judgments, that they inescapably reform or deform (in an ethically significant sense) the putative data of the moral domain itself and that, consequently, one cannot test such theories by any straightforward consultation of independent and antecedently posited moral data. The difficulty is obvious: the data, which should otherwise include standard and correct moral judgments, are, characteristically, adjusted or reinterpreted to fit the very metamoral theories they are to test; alternatively put, the relevant moral data include, on any generous canvassing, the very fact that human agents subscribe to the various competing metamoral principles.

It is hard to see how it can be denied that rational agents attempt to justify their conduct, alternatively, on utilitarian, deontological, and egoistic grounds, that is, that they are convinced—either different agents with respect to the same range of issues or the same agents with respect to different issues—that reasons of the relevant sorts are at least admissible. But if these practices be entered as part of the primary data ethical theories of any sort are concerned to explicate, then, obviously, metaethical theories are bound to be affected by differences in substantive ethical convictions: a utilitarian and a deontologist will not only be unable to agree about the description of the conceptual relationship between “good” and “right” but they will not, in all fairness, be able to see their opponent’s view vindicated by an appeal to the data of the moral domain. Similarly, if the primary data of the moral domain include justificatory appeals to utilitarian, deontological, and egoistic principles (and possibly others), then the prospect of vindicating, in any straightforward way, such principles as the supreme principle for justifying conduct is clearly hopeless. The difficulty, as has been suggested, is that we

⁷ Cf. W. K. Frankena, *Ethics* (Englewood Cliffs, New Jersey, 1963), pp. 16–18.

simply have no way of demarcating the settled data of the moral domain that is clearly neutral and acceptable to the partisans of competing metaethical and metamoral theories. But if this is so, then it is logically impossible to press the relevant disputes in the direction of the truth, and the competing views themselves threaten to reduce to rhetorical maneuvers.

Dispute, of course, is not reduced to rhetoric, but it is, I believe, quite a bit weaker logically than its partisan antagonists are inclined to believe. It is perfectly clear that both metaethical and metamoral theories presuppose some range of moral judgment and moral practice with respect to which they are themselves taken to be confirmed. This, minimally, must include some subset of the actual judgments and practices that a society exhibits and so eliminates utter arbitrariness. If we can agree that, within limits, a society's judgments will include a core recognizable as relatively ineliminable data for any eligible ethical theory, we shall have set at least firm minimal conditions to the testability and defensibility of competing theories. On this view, for instance, the admission of types of cases like those favoring what I have termed egoistic and deontological elements within utilitarianism is tantamount to the defeat of strict utilitarianism as the supreme metamoral principle. It is obvious also that the detailing of such a core is more likely to demonstrate that *no* particular metamoral principle of the sort considered is adequate than to demonstrate that any such particular principle is: the dilemmas that have been mentioned for utilitarianism more convincingly undermine that doctrine than they establish either egoism or deontology. Also, the identification of a minimal core of moral data that metaethical and metamoral theories explicate must be admitted to be, at best, somewhat relativized; for, if our foregoing argument respecting egoism holds, we can hardly expect that egoistic, utilitarian, and deontological theories *could* admit the same data with respect, say, to the practice of punishment. This, of course, suggests a counter-strategy: disputes between competing metamoral theories will be pursued relative to the data admitted; *if* the question concerns, precisely, the adequacy of competing doctrines to accommodate *morally admissible* punishment, egoism will prove to be flatly indefensible and irrelevant.

There is, admittedly, the risk of tendentiousness here, in the concession of initial data; but it cannot be avoided altogether since *some* range of data must be acknowledged as that which the competing

metamoral theories are concerned to explicate. Clearly, viable metamoral debate will have to be occupied with the reasons for which this or that range of questions or this or that range of judgments ought to be recognized as belonging to the core data themselves—where, that is, the issue is *not* to be construed as itself a moral or metamoral issue. An egoism that rules out punishment or duties respecting promise-keeping and contracts as relevant moral questions *simply because* their admission is incompatible with the egoistic principle adopted metamorally is, to that extent, relatively arbitrary and uninteresting—but I cannot see how it can be shown, for that reason merely, to be a false principle. This is, in fact, the chief objection to egoism, and it clarifies considerably the nature of what I am calling metamoral debates. Such debates, to be at all significant, must be addressed to the actual practices and relationships that obtain in a given society; but to reject such debates is not equivalent to holding an indefensible thesis, merely a relatively uninteresting one. Once one admits that metamoral disputes concern putatively overriding values, the importance of this concession cannot be ignored.

If, of course, some form of moral cognitivism could be vindicated, the conceptual difficulties of metamoral and metaethical disputes would not arise at all and a relatively clear division could be provided, *for any particular theoretical dispute*, between *explanandum* and *explanans*. We can only imitate this practice, borrowed from the sciences, in the moral domain—by relativizing particular debates, by conceding that their force is conditional upon the data the competing theories acknowledge *but* where the data cannot be said to be antecedently and independently established. Under the circumstance, therefore, we cannot provide a full model for the testing of metamoral theories analogous with what obtains in the empirical sciences. The imitation of the latter practice is restricted to what might be regarded as a sort of gentleman's agreement: the admission of standard runs of cases thought not to prejudice as such disagreements as between, say, traditional utilitarians and traditional deontologists. Even here, with shifting substantive convictions influenced by adopting the very metamoral principles in question, the imitation will, inevitably, be an extremely fragile one. And once convictions respecting conduct and the justification of conduct, departing rather liberally from the gentleman's agreement noted, be allowed, the fiction of con-

firming and disconfirming metamoral theories will become increasingly obvious.

Under these conditions, I am inclined to think that disputes about metamoral principles may be conducted only in a manner reminiscent of the appreciative disputes among connoisseurs of art. What we have before us is not so much a stable collection of undisputed moral judgments and recognizably defensible moral practices as an evolving tradition of moral debate complete with competing metamoral principles and alternatively systematized fractions of the tradition in terms of such principles. The new metamoral advocate, then, is not so much concerned with a principle in some sense descriptive of the justificatory reasons ultimately offered in the moral domain as he is with a principle by which he may reinterpret appreciatively the entire tradition of morality. Like the connoisseur, he may be expected to be cognizant of the ongoing tradition of judgment consistent

with all factually relevant information as well as capable of illuminating in a fresh way the conceptual connections between portions and details of an enormous domain. But in doing so, he is obviously not bound by the judgments and doctrines of the past, merely bound to begin with them. I take it, for example, that Jesus thought to reform the morality of the Jews by means of a doctrine that gave a new and altered coherence to that morality: I cannot see that disputes about the tenability of those extremely abstract and attenuated theories I have been calling metamoral are in the least degree of a different sort. I take it, therefore, that each in its own turn becomes a part of the ongoing tradition that some new metamoral connoisseur (or reformer) will be bound to accommodate and reinterpret. But if this is so, then some of the most venerable of the disputes of moral philosophy have been, and cannot but have been, seriously misrepresented.

Temple University

Received September 29, 1969

BOOKS RECEIVED

- ADLER, Mortimer J. *The Time of Our Lives: The Ethics of Common Sense*. (New York: Holt, Rinehart and Winston, 1970). Pp. 361. \$7.95
- ARMOUR, Leslie *The Concept of Truth* (The Netherlands: Royal Vangorcum Ltd., 1969). Pp. 254.
- COHEN, Robert S. and WARTOFKY, Marx W. (eds.) *Boston Studies in the Philosophy of Science*. (Dordrecht-Holland: D. Reidel Publishing Company, 1969), vol. 4 (1966/1968). Pp. 537. \$20.00
- *Boston Studies in the Philosophy of Science*. (Dordrecht-Holland: D. Reidel Publishing Company, 1969), vol. 5 (1966/1968). Pp. 482. \$16.75
- ENGEL, S. MORRIS *Language and Illumination: Studies in the History of Philosophy*. (Martinus Nijhoff: The Hague, 1969). Pp. 141.
- ENNIS, Robert H. *Ordinary Logic* (Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1969). Pp. 151. \$2.25 (Paper)
- ESSLER, W. K. *Einführung in die Logik* (Stuttgart, Germany: Alfred Kröner Verlag, 1969). Pp. 325.
- GROSSMAN, Reinhardt *Reflections on Frege's Philosophy* (Evanston, Illinois: Northwestern University Press, 1969). Pp. 261. \$8.75
- MAHADEVAN, T. M. P. *Indian Philosophical Annual*, vol. 2 (1966) (Madras: The University of Madras, 1968). Pp. 342.
- MARTIN, R. M. *Belief, Existence, and Meaning* (New York: New York University Press, 1969). Pp. 284. \$12.50
- MYERS, Gerald E. *Self: An Introduction to Philosophical Psychology* (New York: Pegasus, 1969). Pp. 173. \$1.95 (Paper)
- SCHMIDT, Siegfried J. *Bedeutung und Begriff* (Braunschweig: Friedr. Vieweg & Sohn, 1969). Pp. 176.
- STEGMÜLLER, Wolfgang *Main Currents in Contemporary German, British and American Philosophy* (Dordrecht-Holland: D. Reidel Publishing Company, 1969). Pp. 567.
- Studia Leibnitiana, Suppl. III—Erkenntnislehre* (Wiesbaden, Germany: Franz Steiner Verlag GMBH, 1969). Pp. 242.
- SUPPES, Patrick, *Studies in the Methodology and Foundations of Science: Selected Papers from 1951-1969* (Dordrecht-Holland: D. Reidel Publishing Company, 1969). Pp. 473.
- TEENSMA, E. *The Paradoxes* (The Netherlands: Royal Vangorcum Ltd., 1969). Pp. 44.

DIALOGUE

Canadian Philosophical Review - Revue Canadienne de Philosophie

Editors: VENANT CAUCHY and MARTYN ESTALL

VOL. VIII - 1970 - No. 4

ARTICLES

Interprétations anciennes du fragment 62 d'Héraclite
The Dramatic Aspect of Plato's *Phaedo*
Knowledge and Flux in Plato's *Cratylus* (438-40)
L'Ontologie de Karl Rahner
Peut-on et doit-on fonder la morale
Bultmann's Philosophical Troubles
Berkeley's Ambiguity

JEAN PEPIN
KENNETH DORTER
M. T. THORNTON
ROGER LAPOINTE
BERNARD CARNOIS
H. A. NIELSEN
DAVID A. GIVNER

VOL. IX - 1970 - No. 1

ARTICLES

Analytic and Existential Ethics
Habit and Reflection in Morality
Archéologie du savoir et structures du langage scientifique
La Philosophie médicale de Sydenham
Le Projet blondélien et le souci de l'unité

C. D. MAGNIVEN
G. P. HENDERSON
NORMAND LACHARITE
FRANCOIS DUCHESNEAU
YVON POITRAS

NOTES — DISCUSSIONS — REVIEWS

*Subscriptions: \$10.00 a year to individuals; \$12.00 to libraries.
Payable to the Canadian Philosophical Association in care of Norman J. Brown,
Department of Philosophy, Queen's University, Kingston, Ontario*

The British Journal for the Philosophy of Science

Vol. 21, No. 2, May 1970

R. B. BRAITHWAITE on Bertrand Russell
A. F. CHALMERS Curie's Principle
SCOTT A. KLEINER Erotetic Logic and the Structure of Scientific Revolution
DAVID BLOOR Is the Official Theory of Mind Absurd?
T. G. MCGONIGLE Euclidean Space: A Lasting Philosophical Obsession
LOWELL NISSEN Canfield's Functional Translation Schema
RICHARD J. HALL Kuhn and the Copernican Revolution
G. M. K. HUNT A Conditional Vindication of the Straight Rule
PROFESSOR T. W. SETTLE Confirmation as a Probability: Dead but it won't Lie Down!

CAMBRIDGE UNIVERSITY PRESS

Bentley House, 200 Euston Road, London, N.W.1.
American Branch: 32 East 57th Street, New York, N.Y. 10022
Single issues 20s. net in U.K. (\$3.00 in U.S.A.)
Subscription price 60s. net in U.K. (\$9.50 in U.S.A.)

SYNTHESE

*An International Journal for Epistemology,
Methodology and Philosophy of Science*

Editor-in-Chief: JAAKKO HINTIKKA
University of Helsinki and Stanford University

Board of Consulting Editors: David M. Armstrong, Yehoshua Bar-Hillel, Arthur W. Burks, Robert S. Cohen, Donald Davidson, Hans Freudenthal, Ian Hacking, Béla Juhos, Jerrold J. Katz, Imre Lakatos, G. Nuchelmans, Wesley C. Salmon, Wolfgang Stegmüller, Erik Stenius, Patrick Suppes, Klemens Szaniawski, Ladislav Tondl, H. Törnebohn and Marx Wartofsky.

Contents of Volume 21, No. 2, June 1970:

Risto Hilpinen Knowing That One Knows and the Classical Definition of Knowledge.
Keith Lehrer Believing That One Knows.
Jaakko Hintikka 'Knowing That One Knows' Reviewed.
Carl Ginet What Must Be Added To Knowing To Obtain Knowing That One Knows?
Hector-Neri Castañeda On Knowing (Or Believing) That One Knows (Or Believes).
Peter C. Fishburn Utility Theory With Inexact Preferences and Degrees of Preference.
Myles Brand and Marshall Swain On the Analysis of Causation.
J. Bronowski New Concepts in the Evolution of Complexity. Stratified Stability and Unbounded Plans.

Subscription price per volume of four issues Dfl. 60.—
(U.S. \$17.00)

Personal subscription price Dfl. 35.—(U.S. \$9.80)

D. REIDEL PUBLISHING COMPANY
DORDRECHT — HOLLAND

AMERICAN PHILOSOPHICAL QUARTERLY

MONOGRAPH SERIES

Edited by NICHOLAS RESCHER

The *American Philosophical Quarterly* also publishes a supplementary Monograph Series. Apart from the publication of original scholarly monographs, this is to include occasional collections of original articles on a common theme. The current monograph is included in the subscription, and past monographs can be purchased separately but are available to individual subscribers at *half price* (though not to institutional subscribers).

- No. 1. STUDIES IN MORAL PHILOSOPHY. *Contents:* Kai Nielsen, "On Moral Truth"; Jesse Kalin, "On Ethical Egoism"; G. P. Henderson, "Moral Nihilism"; Michael Stocker, "Supererogation and Duties"; Lawrence Haworth, "Utility and Rights"; David Braybrooke, "Let Needs Diminish That Preferences May Prosper"; and Jerome B. Schneewind, "Whewell's Ethics." 1968, \$6.00.
- No. 2. STUDIES IN LOGICAL THEORY. *Contents:* Montgomery Furth, "Two Types of Denotation"; Jaakko Hintikka, "Language-Games for Quantifiers"; James W. Cornman, "Types, Categories, and Nonsense"; Robert C. Stalnaker, "A Theory of Conditionals"; Alan Hausman and Charles Echelbarger, "Goodman's Nominalism"; Ted Honderich, "Truth: Austin, Strawson, Warnock"; and Colwyn Williamson, "Propositions and Abstract Propositions." 1968, \$6.00.
- No. 3. STUDIES IN THE PHILOSOPHY OF SCIENCE. *Contents:* Peter Achinstein, "Explanation"; Keith Lehrer, "Theoretical Terms and Inductive Inference"; Lawrence Sklar, "The Conventionality of Geometry"; Mario Bunge, "What Are Physical Theories?"; B. R. Grunstra, "The Plausibility of the Entrenchment Concept"; Simon Blackburn, "Goodman's Paradox"; Stephen Spielman, "Assuming, Ascertaining, and Inductive Probability"; Joseph Agassi, "Popper on Learning from Experience"; D. H. Mellor, "Physics and Furniture"; and Michael Slote, "Religion, Science, and the Extraordinary." 1969, \$6.00.
- No. 4. STUDIES IN THE THEORY OF KNOWLEDGE. *Contents:* John Knox, Jr., "Do Appearances Exist?"; Norman Malcolm, "Wittgenstein on the Nature of Mind"; W. Donald Oliver, "A Sober Look at Solipsism"; John L. Pollock, "The Structure of Epistemic Justification"; Frederick Stoutland, "The Logical Connection Argument"; Peter Unger, "Our Knowledge of the Material World"; Alan R. White, "What Might Have Been." 1970, \$6.00.

AMERICAN PHILOSOPHICAL QUARTERLY

Edited by

NICHOLAS RESCHER

With the advice and assistance of the Board of Editorial Consultants:

Virgil C. Aldrich
Alan R. Anderson
Kurt Baier
Stephen F. Barker
Monroe Beardsley
Nuel D. Belnap, Jr.
Roderick M. Chisholm
L. Jonathan Cohen
James Collins
Arthur C. Danto

James M. Edie
José Ferrater-Mora
Richard M. Gale
Peter Thomas Geach
Adolf Grünbaum
Carl G. Hempel
John Hospers
Raymond Klibansky
Hugues Leblanc
Ernan McMullin

Benson Mates
John A. Passmore
Richard H. Popkin
Richard Rorty
George A. Schrader
Michael Scriven
Wilfrid Sellars
Alexander Sesonske
Manley H. Thompson, Jr.
John W. Yolton

VOLUME 7/NUMBER 4

OCTOBER 1970

CONTENTS

I. SYDNEY SHOEMAKER: <i>Persons and Their Pasts</i>	269	VI. DAVID CARR: <i>Husserl's Problematic Concept of the Life-World</i>	331
II. JOHN LESLIE: <i>The Theory That the World Exists Because It Should</i>	286	VII. BENJAMIN GIBBS: <i>Real Possibility</i>	340
III. L. JONATHAN COHEN: <i>Applications of Inductive Logic to Theory of Language</i>	299	VIII. EVAN SIMPSON: <i>Actions and Extensions</i>	349
IV. BERNARD BEROFSKY: <i>Purposive Action</i>	311	IX. JOHN KNOX, JR.: <i>Does Becoming Entail a Contradiction?</i>	357
V. GERALD VISION: <i>Essentialism and the Sense of Proper Names</i>	321	X. EDWARD F. WALTER: <i>Empiricism and Ethical Reasoning</i>	364
		<i>Books Received</i>	371
		<i>Corrigenda</i>	373

AMERICAN PHILOSOPHICAL QUARTERLY

POLICY

The *American Philosophical Quarterly* welcomes articles by philosophers of any country on any aspect of philosophy, substantive or historical. However, only self-sufficient articles will be published, and not news items, book reviews, critical notices, or "discussion notes."

MANUSCRIPTS

Contributions may be as short as 2,000 words or as long as 25,000. All manuscripts should be typewritten with wide margins, and at least double spacing between the lines. Footnotes should be used sparingly and should be numbered consecutively. They should also be typed with wide margins and double spacing. The original copy, not a carbon, should be submitted; authors should always retain at least one copy of their articles.

COMMUNICATIONS

Articles for publication, and all other editorial communications and enquiries, should be addressed to: The Editor, *American Philosophical Quarterly*, Department of Philosophy, University of Pittsburgh, Pittsburgh, Pennsylvania 15213.

OFFPRINTS

Authors will receive gratis two copies of the issue containing their contribution. Offprints can be purchased through arrangements made when checking proof. They will be charged for as follows: The first 50 offprints of 4 pages (or fraction thereof) cost \$12, increasing by \$1 for each additional 4 pages. Additional groups of 50 offprints of 4 pages cost \$8, increasing by \$1 for each additional 4 pages. Covers will be provided for offprints at a cost of \$4 per group of 50.

SUBSCRIPTIONS

The price *per annum* is eight dollars for individual subscribers and fourteen dollars for institutions. Checks and money orders should be made payable to the *American Philosophical Quarterly*. All back issues are available and are sold at the rate of three dollars to individuals, and four dollars to institutions. All correspondence regarding subscriptions and back orders should be addressed directly to the publisher (Basil Blackwell, Broad Street, Oxford, England).

MONOGRAPH SERIES

The *American Philosophical Quarterly* also publishes a supplementary Monograph Series. Apart from the publication of original scholarly monographs, this includes occasional collections of original articles on a common theme. The current monograph is included in the subscription, and past monographs can be purchased separately but are available to individual subscribers at a substantially reduced price. The back cover of the journal may be consulted for details.

335081



I. PERSONS AND THEIR PASTS

SYDNEY SHOEMAKER

PERSONS have, in memory, a special access to facts about their own past histories and their own identities, a kind of access they do not have to the histories and identities of other persons and other things. John Locke thought this special access important enough to warrant a special mention in his definition of "person," viz., "a thinking, intelligent being, that has reason and reflection, and can consider itself as itself, the same thinking thing, in different times and places. . . ."¹ In this paper I shall attempt to explain the nature and status of this special access and to defend Locke's view of its conceptual importance. I shall also attempt to correct what now seem to me to be errors and oversights in my own previous writings on this topic.

I

As a first approximation, the claim that persons have in memory a special access to their own past histories can be expressed in two related claims, both of which will be considerably qualified in the course of this paper. The first is that it is a necessary condition of its being true that a person remembers a given past event that he, that same person, should have observed or experienced the event, or known of it in some other direct way, at the time of its occurrence. I shall refer to this as the "previous awareness condition" for remembering.²

The second claim is that an important class of first person memory claims are in a certain respect

immune to what I shall call "error through misidentification." Consider a case in which I say, on the basis of my memory of a past incident, "I shouted that Johnson should be impeached," and compare this with a case in which I say, again on the basis of my memory of a past incident, "John shouted that Johnson should be impeached." In the latter case it could turn out that I do remember someone who looked and sounded just like John shouting that Johnson should be impeached, but that the man who shouted this was nevertheless not John—it may be that I misidentified the person as John at the time I observed the incident, and that I have preserved this misidentification in memory, or it may be that I subsequently misidentified him as John on the basis of what I (correctly) remembered about him. Here my statement would be false, but its falsity would not be due to a mistake or fault of my memory; my memory could be as accurate and complete as any memory could be without precluding this sort of error. But this sort of misidentification is not possible in the former case. My memory report could of course be mistaken, for one can misremember such incidents, but it could not be the case that I have a full and accurate memory of the past incident but am mistaken in thinking that the person I remember shouting was myself. I shall speak of such memory judgments as being immune to error through misidentification with respect to the first

¹ Locke, *Essay Concerning Human Understanding*, Bk. II, Chap. 27, sec. 9 (London, 1912). Italics added.

² In their paper "Remembering" (*The Philosophical Review*, vol. 75 [April, 1966]) C. B. Martin and Max Deutscher express what I call the previous awareness condition by saying that "a person can be said to remember something happening or, in general, remember something directly, only if he has observed or experienced it." Their notion of direct remembering seems to be much the same as Norman Malcolm's notion of "personal memory" (see his "Three Lectures on Memory" in *Knowledge and Certainty* [Englewood Cliffs, N.J., 1963], pp. 203–221). To remember that Caesar invaded Britain I need not have had any experience of the invasion, but no one who lacked such experience could directly or personally remember that Caesar invaded Britain. In this paper I am primarily concerned with memories that are of events, i.e., of something happening, and do not explicitly consider what Malcolm calls "factual memory," i.e., memories that such and such was (or is, or will be) the case, but what I say can be extended to cover all cases of direct or personal memory. Martin and Deutscher hold, and I agree, that remembering something happening is always direct remembering.

There are apparent counterexamples to the previous witnessing condition as I have formulated it. I can be said to remember Kennedy's assassination, which is presumably an event, yet I did not witness or observe it, and the knowledge I had of it at the time was indirect. But while I can be said to remember the assassination, I could hardly be said to remember Kennedy being shot (what I do remember is hearing about it, and the impact this made on me and those around me). Perhaps I can be said to remember the assassination because we sometimes mean by "the assassination" not only the events in Dallas but their immediate effects throughout the nation and world. In any case, when I speak of memories of events in this paper I mean what Martin and Deutscher speak of as memories of something happening.

person pronouns, or other "self-referring" expressions, contained in them.³

I do not contend that all memory claims are immune to error through misidentification with respect to the first person pronouns contained in them. If I say "I blushed when Jones made that remark" because I remember seeing in a mirror someone, whom I took (or now take) to be myself, blushing, it could turn out that my statement is false, not because my memory is in any way incomplete or inaccurate, but because the person I saw in the mirror was my identical twin or double.⁴ In general, if at some past time I could have known of someone that he was ϕ , and could at the same time have been mistaken in taking that person to be myself, then the subsequent memory claims I make about that past occasion will be subject to error through misidentification with respect to the first person pronouns. But if, as is frequently the case, I could not have been mistaken in this way in the past in asserting what I then knew by saying "I am ϕ ," then my subsequent memory claim "I was ϕ " will be immune to error through misidentification relative to 'I'; that is, it is impossible in such cases that I should accurately remember someone being ϕ but mistakenly take that person to be myself. We might express this by saying that where the present tense version of a judgment is immune to error through misidentification relative to the first person pronouns contained in it, this immunity is *preserved* in memory.⁵ Thus if I claim on the strength of memory that I saw John yesterday, and have a full and accurate memory of the incident, it cannot be the case that I remember someone seeing John but have misidentified that person as myself; my memory claim "I saw John" is subject to error

through misidentification with respect to the term "John" (for it could have been John's twin or double that I saw), but not with respect to 'I'.

II

In his early paper, "Personal Identity," H. P. Grice held that the proposition "One can only remember one's own past experiences" is analytic, but pointed out that this would be analytic in only a trivial way "if 'memory' were to be defined in terms of 'having knowledge of one's own past experiences'." He says that "even if we were to define 'memory' in this sort of way, we should still be left with a question about the proposition, 'one can only have knowledge of one's own past experiences,' which seems to me a necessary proposition."⁶ Now I doubt very much if Grice, or any other philosopher, would now want to hold that is necessarily true, or that it is true at all, that one's own past experiences are the only past experiences of which one can have knowledge. But one does not have to hold this to hold, with Grice, that it is not just a trivial analytic truth that one's own experiences are the only ones that one can remember, i.e., that it is not the case that the necessity of this truth derives merely from the fact that we refuse to *call* someone's having knowledge of a past experience a case of his remembering it unless the past experience belonged to the rememberer himself.

Grice's remarks are explicitly about memory of past experiences, but they raise an important question about all sorts of "event memory." Supposing it to be a necessary truth that the previous witnessing condition must be satisfied in any

³ Although self-reference is typically done with first person pronouns, it can be done with names, and even with definite descriptions—as when De Gaulle says "De Gaulle intends . . .," and the chairman of a meeting says "The Chair recognizes. . . ." In such cases these expressions are "self-referring," not merely because their reference is in fact to the speaker, but also because the speaker intends in using them to refer to himself.

⁴ There is a subtle distinction between this sort of case and cases like the following, which I would not count as a case of error through misidentification. Suppose that Jones says "You are a fool," and I mistakenly think that he is speaking to me. Subsequently I say "I remember Jones calling me a fool," and my statement is false through no fault of my memory. While this is a case of knowing *that* Jones called someone (someone or other) a fool and mistakenly thinking that he was calling me a fool, it is not a case of knowing *of* some particular person that Jones called him a fool but mistakenly identifying that person as oneself. Whereas in the other case we can say, not merely that I know that someone or other blushed, mistakenly think that it was I, but I know *of* some particular person (namely the man I saw in the mirror) that he blushed and have mistakenly identified him as myself.

⁵ I have discussed the immunity to error through misidentification of first person present tense statements in my paper "Self-Reference and Self-Awareness," *The Journal of Philosophy*, vol. 65, 19 (1968). In that paper I made the mistake of associating this feature with the peculiarities of the first-person pronouns. But in fact present tense statements having the appropriate sorts of predicates are immune to error to misidentification with respect to any expressions that are "self-referring" in the sense of footnote 3, including names and definite descriptions. If someone says "De Gaulle intends to remove France from NATO," and is using "De Gaulle" to refer to himself, his statement is in the relevant sense immune to error through misidentification, regardless of whether he is right in thinking his name is "De Gaulle" and that he is the President of France.

⁶ H. P. Grice, "Personal Identity," *Mind*, vol. 50 (1941), p. 344.

genuine case of remembering, is this necessarily true because we would refuse to *count* knowing about a past event as remembering it if the previous awareness condition were not satisfied, or is it necessary for some deeper reason? I think that many philosophers would hold that if this is a necessary truth at all, it is so only in the former way, i.e., in such a way as to make its necessity trivial and uninteresting. Thus G. C. Nerlich, in a footnote to his paper "On Evidence for Identity," says that it is true only of *our* world, not of all possible worlds, that only by being identical with a witness to past events can one have the sort of knowledge of them one has in memory.⁷ On this view it is logically possible that we should have knowledge of past events which we did not ourselves witness, of experiences we did not ourselves have, and of actions we did not ourselves perform, that is in all important respects like the knowledge we have of past events, experiences, and actions in remembering them. If one takes this view it will seem a matter of small importance, if indeed it is true, that the having of such knowledge could not be called "remembering."

It is of course not absolutely clear just what it means to speak of knowledge as being "in all important respects like" memory knowledge, if this is not intended to imply that the knowledge is memory knowledge. Presumably, knowledge of past events that is "just like" memory knowledge must not be inferred from present data (diaries, photographs, rock strata, etc.) on the basis of empirical laws and generalizations. But while this is necessary, it is not sufficient. When a person remembers a past event there is a correspondence between his present cognitive state and some past cognitive and sensory state of his that existed at the time of the remembered event and consisted in his experiencing the event or otherwise being aware of

its occurrence.⁸ I shall say that remembering a past event involves there being a correspondence between the rememberer's present cognitive state and a past cognitive and sensory state that was "of" the event.⁹ In actual memory this past cognitive and sensory state is always a past state of the rememberer himself. What we need to consider is whether there could be a kind of knowledge of past events such that someone's having this sort of knowledge of an event does involve there being a correspondence between his present cognitive state and a past cognitive and sensory state that was of the event, but such that this correspondence, although otherwise just like that which exists in memory, does not necessarily involve that past state's having been a state of the very same person who subsequently has the knowledge. Let us speak of such knowledge, supposing for the moment that it is possible, as "quasi-memory knowledge," and let us say that a person who has this sort of knowledge of a past event "quasi-remembers" that past event. Quasi-remembering, as I shall use the term, includes remembering as a special case. One way of characterizing the difference between quasi-remembering and remembering is by saying that the former is subject to a weaker previous awareness condition than the latter. Whereas someone's claim to remember a past event implies that he himself was aware of the event at the time of its occurrence, the claim to quasi-remember a past event implies only that someone or other was aware of it. Except when I indicate otherwise, I shall use the expression "previous awareness condition" to refer to the stronger of these conditions.

Our faculty of memory constitutes our most direct access to the past, and this means, given the previous awareness condition, that our most direct access to the past is in the first instance an access to *our own* past histories. One of the main questions I

⁷ G. C. Nerlich, "On Evidence for Identity," *Australian Journal of Philosophy*, vol. 37 (1959), p. 208.

⁸ I am not here endorsing the view, which I in fact reject, that remembering consists in the having of an image, or some other sort of mental "representation," in which the memory content is in some way encoded. It is sufficient for the existence at *t* of the "cognitive state" of remembering such and such that it be true of the person at *t* that he remembers such and such; I am not here committing myself to any account of what, if anything, someone's remembering such and such "consists in."

⁹ I should make it clear that I am not saying that what we remember is always, or even normally, a past cognitive and sensory state. I am not propounding the view, which is clearly false, that "strictly speaking" one can remember only one's own past experiences. I am saying only that if a person remembers an event that occurred at time *t* then at *t* there must have been a corresponding cognitive and sensory state—which the person may or may not remember—that was of that event. It would not be easy to specify just what sort of correspondence is required here, and I shall not attempt to do so. But I take it as obvious that the claim to remember firing a gun requires, for its truth, a different sort of past cognitive and sensory state than the claim to remember hearing someone else fire a gun, and that the latter, in turn, requires a different sort of past cognitive and sensory state than the claim to remember seeing someone fire a gun. Sometimes one remembers a past event but no longer remembers just how one knew of it at the time of its occurrence; in such a case one's memory, because of vagueness and incompleteness, corresponds to a wider range of possible cognitive and sensory states than (say) a memory of seeing the event or a memory of being told about it.

shall be considering in this paper is whether it is conceivable that our most direct access to the past should be a faculty of quasi-remembering which is not a faculty of remembering. Is it conceivable that we should have, as a matter of course, knowledge that is related to past experiences and actions other than our own in just the way in which, as things are, our memory knowledge is related to our own past experiences and actions? In our world all quasi-remembering is remembering; what we must consider is whether the world could be such that most quasi-remembering is not remembering.

Before going on to consider this question I should mention two reasons why I think it important. The first is its obvious bearing on the question of the relationship between the concepts of memory and personal identity. If there can be quasi-remembering that is not remembering, and if remembering can be defined as quasi-remembering that is of events the quasi-rememberer was aware of at the time of their occurrence (thus making it a trivial analytic truth that one can remember an event only if one was previously aware of it), then it would seem that any attempt to define or analyze the notion of personal identity in terms of the notion of remembering will be viciously circular. I shall have more to say about this in Sect. V. But this question also has an important bearing on the question of how a person's memory claims concerning his own past are grounded. In previous writings I have claimed, and made a great deal of the claim, that our memory knowledge of our own past histories, unlike our knowledge of the past histories of other things, is not grounded on criteria of identity.¹⁰ Strawson makes a similar claim in *The Bounds of Sense*, saying that "When a man (a subject of experience) ascribes a current or directly remembered state of consciousness to himself, no use whatever of any criteria of personal identity is required to justify his use of the pronoun 'I' to refer to the subject of that experience." He remarks that "it is because Kant recognized this truth that his treatment of the subject is so greatly superior to Hume's."¹¹ Now it can easily seem that this claim

follows immediately from the fact that remembering necessarily involves the satisfaction of the previous awareness condition. If one remembers a past experience then it has to have been one's own, and from this it may seem to follow that it makes no sense to inquire concerning a remembered experience whether it was one's own and then to try to answer this question on the basis of empirical criteria of identity. But suppose that it were only a trivial analytic truth that remembering involves the satisfaction of the previous awareness condition, and suppose that it were possible to quasi-remember experiences other than one's own. If this were so one might remember a past experience but not know whether one was remembering it or only quasi-remembering it. Here, it seems, it would be perfectly appropriate to employ a criterion of identity to determine whether the quasi-remembered experience was one's own, i.e., whether one remembered it as opposed to merely quasi-remembering it. Thus the question of whether the knowledge of our own identities provided us by memory is essentially non-critical turns on the question of whether it is possible to quasi-remember past actions and experiences without remembering them.

III

There is an important respect in which my characterization of quasi-remembering leaves that notion inadequately specified. Until now I have been ignoring the fact that a claim to remember a past event implies, not merely that the rememberer experienced such an event, but that his present memory is in some way *due to*, that it came about *because of*, a cognitive and sensory state the rememberer had at the time he experienced the event. I am going to assume, although this is controversial, that it is part of the previous awareness condition for memory that a veridical memory must not only correspond to, but must also stand in an appropriate *causal* relationship to, a past cognitive and sensory state of the rememberer.¹² It may seem that if quasi-memory is to

¹⁰ See my book *Self-Knowledge and Self-Identity* (Ithaca, N.Y., 1963), especially Chap. Four, and my paper "Personal Identity and Memory," *Journal of Philosophy*, vol. 56 (1959), pp. 868-882.

¹¹ P. F. Strawson, *The Bounds of Sense* (London, 1966), p. 165.

¹² I owe to Norman Malcolm the point that to be memory knowledge one's knowledge must be in some way due to, must exist because of, a past cognitive and sensory state of oneself—see his "Three Lectures on Memory" (*op. cit.*). Malcolm holds that "due to" does not here express a causal relationship, but I have been persuaded otherwise by Martin's and Deutscher's "Remembering" (*op. cit.*). See also my paper "On Knowing Who One Is" (*Common Factor*, No. 4, 1966), and David Wiggins' *Identity and Space-Time Continuity* (Oxford, 1967), especially p. 50 ff. The view that there is a causal element in the concept of memory is attacked by Roger Squires in his recent paper "Memory Unchained" (*The Philosophical Review*, vol. 78 [1969] pp. 178-196); I make a very limited reply to this in Sect. V of this paper.

be as much like memory as possible, we should build a similar requirement into the previous awareness condition for quasi-memory, i.e., that we should require that a veridical quasi-memory must not only correspond to, but must also stand in an appropriate causal relationship to, a past cognitive and sensory state of someone or other. On the other hand, it is not immediately obvious that building such a requirement into the previous awareness condition for quasi-memory would not make it equivalent to the previous awareness condition for memory, and thus destroy the intended difference between memory and quasi-memory. But there is no need for us to choose between a previous awareness condition that includes the causal requirement and one that does not, for it is possible and useful to consider both. In the present section I shall assume that the previous awareness condition for quasi-memory does not include the causal requirement, and that it includes nothing more than the requirement that a quasi-memory must, to be a veridical quasi-memory of a given event, correspond in content to a past cognitive and sensory state that was of that event. In the sections that follow I shall consider the consequences of strengthening this condition to include the causal requirement.

The first thing we must consider is what becomes of the immunity of first person memory claims to error through misidentification if we imagine the faculty of memory replaced by a faculty of quasi-memory. As things are now, there is a difference between, on the one hand, remembering an action of someone else's—this might consist, for example, in having a memory of seeing someone do the action—and, on the other hand, remembering *doing* an action, which can be equated with remembering *oneself* doing the action. In the case of quasi-remembering the distinction corresponding to this is that between, on the one hand, the sort of quasi-memory of a past action whose corresponding past cognitive and sensory state belonged to someone who was watching someone else do the action, and, on the other hand, the sort of quasi-memory of a past action whose corresponding past cognitive and sensory state belonged to the very person who did the action. Let us call these, respectively, quasi-memories of an action "from the outside" and quasi-memories of an action "from the inside." Now whereas I can remember an action from the inside only if it was my action, a world in which there is quasi-remembering that is not remembering will be one in which it is not true that any

action one quasi-remembers from the inside is thereby an action he himself did. So—assuming that ours may be such a world—if I quasi-remember an action from the inside, and say on this basis that I did the action, my statement will be subject to error through misidentification; it may be that my quasi-memory of the action is as accurate and complete as it could be, but that I am mistaken in thinking that I am the person who did it. There is another way in which a first person quasi-memory claim could be mistaken through misidentification. If there can be quasi-remembering that is not remembering, it will be possible for a person to quasi-remember an action of his own from the outside. That is, one might quasi-remember an action of one's own as it appeared to someone else who observed it; one might, as it were, quasi-remember it through the eyes of another person. But of course, if I were to quasi-remember someone who looks like me doing a certain action, and were to say on that basis that I did the action, I might be mistaken through no fault of my quasi-memory; it might be that the person who did the action was my identical twin or someone disguised to look like me.

What I have just said about the quasi-remembering of past actions also applies to the quasi-remembering of past experiences and of other mental phenomena. If I remember a past pain from the inside—i.e., remember the pain itself, or remember having the pain, as opposed to remembering seeing someone manifest pain behavior—then the pain must have been mine. But the fact that I *quasi*-remember a pain from the inside will be no guarantee that the pain was mine. Any quasi-memory claim to have been in pain on some past occasion, or to have had a certain thought, or to have made a certain decision, will be subject to error through misidentification.

What is shown by the foregoing is that the immunity of first person memory claims to error through misidentification exists only because remembering requires the satisfaction of the previous awareness condition, and that this feature disappears once we imagine this requirement dropped. Quasi-memory, unlike memory, does not preserve immunity to error through misidentification relative to the first person pronouns. To consider the further consequences of replacing memory with quasi-memory, I must first say something more about memory.

To refer to an event of a certain sort as one that one remembers does not always uniquely identify

it, since one may remember more than one event of a given sort, but it does go some way toward identifying it. In referring to an event in this way one to a certain extent locates it in space and time, even if the description of the event contains no place-names, no names of objects by reference to which places can be identified, and no dates or other temporal indicators. For in saying that one remembers the event one locates it within a spatio-temporal region which is defined by one's own personal history. The spatiotemporal region which is "rememberable" by a given person can be charted by specifying the intervals of past time during which the person was conscious and by specifying the person's spatial location, and indicating what portions of his environment he was in a position to witness, at each moment during these intervals. If someone reports that he remembers an event of a certain kind, we know that unless his memory is mistaken an event of that kind occurred within the spatiotemporal region rememberable by him, and in principle we can chart this region by tracing his history back to its beginning.

Ordinarily, of course, we have far more knowledge than this of the spatiotemporal location of a remembered event, for usually a memory report will fix this position by means of dates, place-names, and other spatial and temporal indicators. But it must be noted that memory claims are subject to error through misidentification with respect to spatial indicators. If a man says "I remember an explosion occurring right in front of that building," it is possible for this to be false even if the memory it expresses is accurate and detailed; the remembered explosion may have occurred, not in front of the building indicated, but in front of another building exactly like it. This remains true no matter how

elaborate and detailed we imagine the memory claim to be. For any set of objects that has actually existed in the world, even if this be as extensive as the set of buildings, streets, parks, bridges, etc., that presently make up New York City, it is logically possible that there should somewhere exist, or that there should somewhere and at some time have existed, a numerically different but exactly similar set of objects arranged in exactly the same way. So memory claims are, in principle, subject to error through misidentification even with respect to such place names as "New York City." Here I am appealing to what Strawson has referred to as the possibility of "massive reduplication."¹³

When a memory report attempts to fix the location of a remembered event by reference to some landmark, we are ordinarily justified in not regarding it as a real possibility that the claim involves error through misidentification owing to the reduplication of that landmark. Certainly we are so justified if the landmark is New York City. But it is important to see why this is so. It is not that we have established that nowhere and at no time has there existed another city exactly like New York; as a self-consistent, unrestricted, negative existential claim, this is something that it would be impossible in principle for us to establish.¹⁴ What we can and do know is that New York is not reduplicated within any spatiotemporal region of which anyone with whom we converse can have had experience. Whether or not New York is reduplicated in some remote galaxy or at some remote time in the past, we know that the man who claims to remember doing or experiencing something in a New York-like city cannot have been in any such duplicate. And from this we can conclude that if he does remember doing or ex-

¹³ P. F. Strawson, *Individuals* (London 1959), p. 20.

¹⁴ It will perhaps be objected that the dictum that unrestricted negative existential claims are unfalsifiable in principle is brought into question by the possibility that we might discover—what some cosmologists hold there is good reason for believing—that space and past time are finite. If we discovered this, why shouldn't we be able, at least in principle, to establish that at no place does there exist, and at no time in the past has there existed, a duplicate of New York?

One way of countering this objection would be to introduce the possibility, which has been argued by Anthony Quinton in his paper "Spaces and Times" (*Philosophy*, vol. 57 [1962] pp. 130-141), of there being a multiplicity of different and spatially unrelated spaces. Establishing that there is no duplicate of New York in our space would not establish that there is no space in which there is such a duplicate, and if it is possible for there to be multiplicity of spaces there would seem to be no way in which the latter could be established.

But we needn't have recourse to such recondite possibilities in order to counter the objection, if it is viewed as an objection to my claim that it is the fact that remembering involves the satisfaction of the previous awareness condition that makes it possible for us to rule out the possibility that memory claims are false through misidentification owing to the reduplication of landmarks. For to discover that space or past time is finite, and that massive reduplication does not occur, one would have to have a vast amount of empirical information about the world, including information about the histories of particular things. But, as I think the remainder of my discussion should make clear, one could not be provided with such information by memory (or by quasi-memory) unless one were *already* entitled in a large number of cases to refer to particular places and things in one's memory reports without having to regard it as possible that one's references were mistaken owing to massive reduplication. So this entitlement would have to precede the discovery that space and past time are finite, and could not depend on it.

periencing something in a New York-like city, then it was indeed in New York, and not in any duplicate of it, that the remembered action or event occurred. But we can conclude this only because remembering involves the satisfaction of the previous awareness condition.

Even when a landmark referred to in someone's memory claim is reduplicated within the spatiotemporal region rememberable by that person, we can often be confident that the claim does not involve error through misidentification. Suppose that someone locates a remembered event, say an explosion, by saying that it occurred in front of his house, and we know that there are many houses, some of which he has seen, that are exactly like his. If he reported that he had simply found himself in front of his house, with no recollection of how he had gotten there, and that after seeing the explosion he had passed out and awakened later in a hospital, we would think it quite possible that he had misidentified the place at which the remembered explosion occurred. But suppose instead that he reports that he remembers walking home from work, seeing the explosion in front of his house, and then going inside and being greeted by his family. Here a misidentification of the place of the explosion would require the reduplication, not merely of his house, but also of his family, his place of work, and the route he follows in walking home from work. We could know that no such reduplication exists within the spatiotemporal region of which he has had experience, and could conclude that his report did not involve an error through misidentification. But again, what would enable us to conclude this is the fact that remembering involves the satisfaction of the previous awareness condition.

Presumably, what justifies any of us in using such expressions as "New York" and "my house" in his own memory reports are considerations of the same kind as those that justify others in ruling out the possibility that claims containing such expressions involve error through misidentification. What justifies one is the knowledge that certain sorts of reduplication do not in fact occur within the spatiotemporal regions of which any of us have had experience. Normally no such justification is needed for the use of 'I' in memory reports; this is what is involved in saying that memory claims are normally immune to error through misidentification relative to the first person pronouns. But what makes such a justification possible in the case of "New York" is the same as what makes it un-

necessary in the case of 'I', namely the fact that remembering involves the satisfaction of the previous awareness condition. So it is because of this fact that remembering can provide us, not merely with the information that an event of a certain sort has occurred somewhere or other in the vicinity of persons and things satisfying certain general descriptions, but with the information that such an event occurred in a certain specified place, in a certain specifiable spatial relationship to events presently observed, and in the vicinity of certain specified persons or things. But this is also to say that it is this fact about remembering that makes it possible for us to know that an object or person to which one remembers something happening is, or is not, identical with an object or person presently observed. And it will emerge later that it is also this fact about remembering that makes it possible to know that different memories are, or are not, of events in the history of a single object or person.

But now let us consider the consequences of replacing the faculty of memory by a faculty of quasi-memory. Quasi-remembering does not necessarily involve the satisfaction of the previous awareness condition, and first person quasi-memory claims are, as we have seen, subject to error through misidentification. It is a consequence of this that even if we are given that someone's faculty of quasi-memory is highly reliable, in the sense that when he seems to quasi-remember an event of a certain sort he almost always does quasi-remember such an event, nevertheless his quasi-memory will provide neither him nor us with any positive information concerning the spatial location of the events he quasi-remembers, or with any information concerning the identity, or concerning the history, of any object or person to which he quasi-remembers something happening. The fact that he quasi-remembers an event of a certain sort will not provide us with the information that such an event has occurred within the spatiotemporal region of which he has had experience. But in consequence of this, if he attempts to locate the quasi-remembered event by reference to some object or place known to us, e.g., New York or Mt. Everest, it is impossible for us to rule out on empirical grounds the possibility that his claim involves error through misidentification owing to the reduplication of that object or place. To rule this out we would have to have adequate grounds for asserting, not merely that there is no duplicate of New York (say) in the spatiotemporal region of

which he has had experience, but that at no place and time has there been a duplicate of New York. And this we could not have.¹⁵ But this means that in expressing his quasi-memories he could not be justified in using such expressions as "New York" and "Mt. Everest," or such expressions as 'I', "this," and "here," to refer to the places, persons, and things in or to which he quasi-remembers certain things happening. The most he could be entitled to assert on the basis of his quasi-memories would be a set of general propositions of the form "An event of type ϕ at some time occurred in the history of an object of type A while it stood in relations $R_1, R_2, R_3 \dots$ to objects of types $B, C, D \dots$ " And given only a set of propositions of this sort, no matter how extensive, one could not even begin to reconstruct any part of the history of the world; one could not even have grounds for asserting that an object mentioned in one proposition in the set was one and the same as an object mentioned in another proposition of the set.

So far I have been ignoring the fact that the events and actions we remember generally have temporal duration, and the fact that we sometimes remember connected sequences of events and actions lasting considerable lengths of time. What will correspond to this if remembering is replaced with quasi-remembering? If someone says "I remember doing X and then doing Y ," it would make no sense to say to him, "Granted that your memory is accurate, and that such a sequence of actions did occur, are you sure that it was one and the same person who did both X and Y ?" But now suppose that someone says "I quasi-remember doing X and then doing Y ," and that the world is such that there is quasi-remembering that is not remembering. Here it is compatible with the accuracy of the man's quasi-memory that he should be mistaken in thinking that he himself did X and Y . And as I shall now try to show, it must also be compatible with the accuracy of this man's quasi-memories that he should be mistaken in thinking even that one and the same person did both X and Y .

Suppose that at time t_1 a person, call him A , does action Y and has while doing it a quasi-memory from the inside of the immediately previous occurrence of the doing of action X . A 's having this quasi-memory of the doing of X is of course compatible with X 's having been done by some-

one other than himself. At t_1 A 's cognitive state includes this quasi-memory from the inside of the doing of X together with knowledge from the inside of the doing of Y ; we might say that it includes knowledge from the inside of the action sequence X -followed-by- Y . But now suppose that at a later time t_2 someone, call him B , has a quasi-memory corresponding to the cognitive state of A at t_1 . It would seem that B 's quasi-memory will be a quasi-memory from the inside of the action sequence X -followed-by- Y . This quasi-memory will be veridical in the sense that it corresponds to a past cognitive state that was itself a state of knowledge, yet its being veridical in this way is compatible with X and Y having been done by different persons. If A were mistakenly to assert at t_1 that X and Y were done by the same person, his mistake would not be due to a faulty quasi-memory. And if B 's cognitive state at t_2 corresponds to A 's cognitive state at t_1 , then if B were mistaken at t_2 in thinking that X and Y were done by the same person, this mistake would not be due to a faulty quasi-memory.

If, as I have been arguing, someone's quasi-remembering from the inside the *action* sequence X -followed-by- Y provides no guarantee that X and Y were done by the same person, then by the same reasoning someone's quasi-remembering the *event* sequence X -followed-by- Y provides no guarantee that X and Y were witnessed by the same person, and therefore no guarantee that they occurred in spatial proximity to one another. But any temporally extended event can be thought of as a succession of temporally and spatially contiguous events; e.g., a stone's rolling down a hill can be thought of as consisting in its rolling half of the way down followed by its rolling the other half of the way. Suppose, then, that someone has a quasi-memory of the following event sequence: stone rolling from top of hill to middle followed by stone rolling from middle of hill to bottom. If we knew this to be a memory, and not just a quasi-memory, we would know that if it is veridical then one and the same person observed both of these events, one immediately after the other, and this together with the contents of the memory could guarantee that one and the same hill and one and the same stone were involved in both, and that a single stone had indeed rolled all the way down a hill. But the veridicality of this quasi-memory *qua* quasi-memory would be compatible with these events

¹⁵ The point made in the preceding footnote can now be expressed by saying that even if we, who have the faculty of memory, could establish that at no place and time has there been a duplicate of New York, this could not be established by someone whose faculty of knowing the past was a faculty of quasi-memory.

having been observed by different persons, and with their involving different stones and different hills; it would be compatible with no stone's having rolled all of the way down any hill. And since any temporally extended event can be thought of as a succession of temporally and spatially contiguous events, it follows that someone's quasi-remembering what is ostensibly a temporally extended event of a certain kind is always compatible with there actually being no such event that he quasi-remembers, for it is compatible with his quasi-memory being, as it were, compounded out of quasi-memories of a number of different events that were causally unrelated and spatiotemporally remote from one another. The knowledge of the past provided by such a faculty of quasi-memory would be minimal indeed.¹⁶

IV

But now we must consider the consequences of strengthening the previous awareness condition for quasi-remembering to include the requirement that a veridical quasi-memory must not only correspond to, but must also stand in an appropriate causal relationship to, a past cognitive and sensory state of someone or other. Clearly, much of what I have said about quasi-remembering ceases to hold once its previous awareness condition is strengthened in this way. If, as is commonly supposed, causal chains must be spatiotemporally continuous, then if quasi-memory claims implied the satisfaction of this strengthened previous

awareness condition they would, when true, provide some information concerning the location of the quasi-remembered events and actions. We would know at a minimum that the spatiotemporal relationship between the quasi-remembered event and the making of the quasi-memory claim is such that it is possible for them to be linked by a spatiotemporally continuous causal chain, and if we could trace the causal ancestry of the quasi-memory we could determine precisely when and where the quasi-remembered event occurred. Thus if we construe the previous awareness condition of quasi-memory as including this causal requirement, it seems that a faculty of quasi-remembering could enable us to identify past events and to reidentify persons and things, and it seems at first glance (though not, I think, on closer examination) that it would enable us to do this without giving us a special access to our own past histories.

It must be stressed that this strengthened previous awareness condition is an improvement on the weaker one *only* on the assumption that causal chains (or at any rate the causal chains that link cognitive and sensory states with subsequent quasi-memories) must be spatiotemporally continuous, or at least must satisfy a condition similar to spatiotemporal continuity. If the sort of causality operating here allowed for action at a spatial or temporal distance, and if there were no limit on the size of the spatial or temporal gaps that could exist in a causal chain linking a cognitive and sensory state with a subsequent quasi-memory, then the claim that a quasi-memory originated

¹⁶ It may be objected that I have overlooked one way in which a quasi-rememberer might begin to reconstruct his own past history, and the histories of other things, from the information provided him by his quasi-memories. The quasi-rememberer's difficulties would be solved if he had a way of sorting out those of his quasi-memories that are of his own past, i.e., are memories, from those that are not. But it may seem that the quasi-rememberer could easily tell which of his quasi-memories of the very *recent* past are of his own past, namely by noting which of them have contents very similar to the contents of his *present* experiences; e.g., if he quasi-remembers from the inside the very recent seeing of a scene that resembles very closely the scene he presently sees, it may seem that he can justifiably conclude that the quasi-remembered seeing was his own. And it may seem that by starting in this way he could trace back his own history by finding among his quasi-memories a subset of situations that form a spatiotemporally continuous series of situations, that series terminating in the situation he presently perceives.

This objection assumes that the quasi-rememberer can know the degree of recentness of the situations of which he has quasi-memories, but I shall not here question this assumption. What I shall question is the assumption that if the quasi-rememberer knows that a quasi-remembered scene occurred only a moment or so ago, and that it closely resembles the scene he presently sees, he is entitled to believe that it is numerically the same scene as the one he presently sees and that in all probability it was he who saw it. For of course it could be the case that there is somewhere else a duplicate of the scene he sees, and that his quasi-memory is of that duplicate. It will perhaps be objected that while this is logically possible (given the possibility of quasi-remembering that is not remembering), it is highly improbable. But while it may be intrinsically improbable that a highly complicated situation should be reduplicated within some limited spatiotemporal area, it does not seem intrinsically improbable that such a situation should be reduplicated somewhere or other in the universe—unless the universe is finite, which is something the quasi-rememberer could have no reason for believing (see footnotes 14 and 15). Moreover, one could not be in a position to know how rare or frequent such reduplication is in fact, and therefore how likely or unlikely it is that a given situation is reduplicated, unless one already had a way of reidentifying places and things. So the quasi-rememberer could not be in a position to know this, for he could have a way of reidentifying places and things only if he were already in a position to rule out reduplication as improbable.

in a corresponding cognitive and sensory state would be as unfalsifiable, and as uninformative, as the claim that it corresponds to a past cognitive and sensory state of someone or other.

To consider the consequences of strengthening the previous awareness condition for quasi-memory in the way just suggested I shall have to introduce a few technical expressions. First, I shall use the expressions "quasi_c-remember" and "quasi_c-memory" when speaking of the sort of quasi-remembering whose previous awareness condition includes the causal requirement. Second, I shall use the term "M-type causal chain" to refer to the sort of causal chain that must link a quasi_c-memory with a corresponding past cognitive and sensory state if they are to be "of" the same event, or if the former is to be "of" the latter. Since quasi_c-remembering is to be as much like remembering as is compatible with the failure of the strong previous awareness condition, M-type causal chains should resemble as much as possible the causal chains that are responsible for actual remembering, i.e., should resemble them as much as is compatible with their sometimes linking mental states belonging to different persons. At any given time a person can be said to have a total mental state which includes his memories or quasi_c-memories and whatever other mental states the person has at that time. Let us say that two total mental states, existing at different times, are directly M-connected if the later of them contains a quasi_c-memory which is linked by an M-type causal chain to a corresponding cognitive and sensory state contained in the earlier. And let us say, by way of giving a recursive definition, that

two total mental states are M-connected if either (1) they are directly M-connected, or (2) there is some third total mental state to which each of them is M-connected.¹⁷

Now there are two cases we must consider. Either the world will be such, or it will not, that a total mental state existing at a particular time can be M-connected with at most one total mental state existing at each other moment in time. Or, what comes to the same thing, either the world will be such, or it will not, that no two total mental states existing at the same time can be M-connected. Let us begin by considering the case in which the former of these alternatives holds. This is the case that will exist if there is no "branching" of M-type causal chains, i.e., if it never happens that an M-type causal chain branches into two such chains which then produce quasi_c-memories belonging to different and simultaneously existing total mental states, and if it never happens that different M-type causal chains coalesce and produce in a single total mental state quasi_c-memories whose corresponding past cognitive and sensory states belonged to different and simultaneously existing total mental states. This is presumably the situation that exists in the actual world. And I think that in any world in which this situation exists M-connected total mental states will be, to use a term of Bertrand Russell's, "copersonal," i.e., states of one and the same person, and quasi_c-remembering will reduce to remembering. There seems to me to be at least this much truth in the claim that memory is constitutive of personal identity.¹⁸ (But more about this in Sect. V.)

Now let us consider the case in which M-type

¹⁷ It is worth mentioning that if quasi_c-remembering is to be as much like remembering as possible then not just any causal chain linking a past cognitive and sensory state with a subsequent quasi_c-memory can be allowed to count as an M-type causal chain. For as Martin and Deutscher (*op. cit.*) point out, there are various sorts of cases in which a man's knowledge of a past event is causally due to his previous experience of it but in which the causal connection is obviously not of the right kind to permit us to say that he remembers the event. E.g., I have completely forgotten the event, but know of it now because you told me about it, and you came to know about it through my telling you about it prior to my forgetting it. It is easier to decide in particular cases whether the causal connection is "of the right kind" than it is to give a general account of what it is for the causal connection to be of the right kind, i.e., what it is for there to be an M-type causal chain. I shall not attempt to do the latter here. The notion of an M-type causal chain would of course be completely useless if it were impossible to determine in any particular case whether the causal connection is "of the right kind" without already having determined that the case is one of remembering—but I shall argue in Sect. V that this is not impossible.

¹⁸ In his paper "Bodily Continuity and Personal Identity: A Reply" (*Analysis*, vol. 21 [1960] pp. 42-48), B. A. O. Williams says that "identity is a one-one relation, and . . . no principle can be a criterion of identity for things of type *T* if it relies on what is logically a one-many or many-many relation between things of type *T*," and remarks that the relation "being disposed to make sincere memory claims which exactly fit the life of" is a many-one relation and "hence cannot possibly be adequate in logic to constitute a criterion of identity" (pp. 44-45). Now it may seem that my version of the view that memory is a criterion of personal identity is open to the same objection, for if M-type causal chains can branch and coalesce then the relation "has a quasi-memory which is linked by an M-type causal chain with a cognitive and sensory state of" is not logically a one-one relation. But while this relationship is not logically one-one, the relationship "has a quasi-memory which is linked by a *non-branching* M-type causal chain with a cognitive and sensory state of" is logically one-one, and it is the holding of the latter relationship that I would hold to be a criterion, in the sense of being a sufficient condition, for personal identity.

causal chains do sometimes branch, and in which, as a result, it can happen that two or more simultaneously existing total mental states are M-connected. Here we cannot claim that if two total mental states are M-connected they are thereby copersonal without committing ourselves to the unattractive conclusion that a person can be in two different places, and can have two different total mental states, at one and the same time. But it is still open to us to say that if a total mental state existing at time t_1 and a total mental state existing at time t_2 are M-connected then they are copersonal *unless* the M-type causal chain connecting them branched at some time during the interval t_1 – t_2 . If we can say this, as I think we can, then even in a world in which there is branching of M-type causal chains the fact that a person quasi_c-remembers a past event or action would create a presumption that he, that same person, experienced the event or did the action, and therefore a presumption that the quasi_c-memory was actually a memory. This presumption would stand as long as there was no evidence that the M-type causal chain linking the past action or experience with the subsequent quasi_c-memory had branched during the interval between them.

Worlds of the sort we are now considering, i.e., worlds in which M-type causal chains sometimes branch, could be of several kinds. Consider first a world in which people occasionally undergo fission or fusion; i.e., people sometimes split, like amoebas, both offshoots having quasi_c-memories of the actions done prior to the fission by the person who underwent it, and two people sometimes coalesce into a single person who then has quasi_c-memories of both of their past histories. Here we cannot say that a person did whatever actions he quasi_c-remembers from the inside without running afoul of Leibniz' Law and the principle of the transitivity of identity. But we can say something close to this. Suppose that someone, call him Jones, splits into two persons, one of whom is I and the other is

someone I shall call Jones' II. Both Jones I and II have quasi_c-memories from the inside of Jones's past actions, and no one else does. If anyone now alive is identical with Jones it is either myself or Jones II, and any objection to saying that I am Jones is equally an objection to saying that Jones II is Jones. I think that we can say here that I am identical with Jones if anyone now alive is identical with him. Or suppose that two people, call them Brown and Smith, coalesce, resulting in me. I have quasi_c-memories from the inside of Brown's actions and also of Smith's actions. There are serious objections to identifying me with either Brown or Smith, but it seems clear here that if anyone now alive is identical with either Brown or Smith, I am. So in such a world the following principle holds: if at time t a person A quasi_c-remembers a past action X from the inside then A is identical with the person who did X if anyone alive at t is identical with him.¹⁹

But I think that we can imagine a world in which this principle would not hold. In the case in which two persons coalesce the M-type causal chains involved might be represented by a river having two "forks" of equal width. Suppose that instead of this we have an M-type causal chain, or a connected set of such causal chains, that could be represented by a river having several small tributaries. For example, suppose, very fancifully, that memories were stored, by some sort of chemical coding, in the blood rather than in brain cells, and that as a result of being given a blood transfusion one sometimes acquired quasi_c-memories "from the inside" of a few of the actions of the blood donor. Here the blood transfusion would be a "tributary" into what apart from its tributaries would be the sort of M-type causal chain that occurs in the history of a single person. Now I do not think that we would deny that A , existing at time t_2 was the same person as B , who existed at an earlier time t_1 , merely because A quasi_c-remembers from the inside, as the result of a blood transfusion, an

¹⁹ A. N. Prior has defended the view that in cases of fission *both* offshoots can be identified with the original person, although not with each other. This of course involves modifying the usual account of the logical features of identity. See his "Opposite Number" (*Review of Metaphysics*, vol. 11 [1957] pp. 196–201), and his "Time, Existence and Identity" (*Proceedings of the Aristotelian Society*, 1955–1966). Roderick Chisholm takes a very different view. Considering the supposition that "you knew that your body, like that of an amoeba, would one day undergo fission and that you would go off, so to speak, in two different directions," he says "it seems to me, first, that there is no possibility whatever that *you* would be *both* the person on the right and the person on the left. It seems to me, secondly, that there is a possibility that you would be one or the other of those two persons" ("The Loose and Popular and the Strict and Philosophical Senses of Identity," in *Perception and Personal Identity*, ed. by Norman S. Care and Robert H. Grimm [Cleveland, 1969], p. 106). It is not clear to me whether Chisholm would hold that one (but not both) of the offshoots might be me if the memories of each stood in the same causal relationships to my actions and experiences as the memories of the other, and if each resembled me, in personality, appearance, etc., as much as the other. If so, I would disagree.

action at t_1 that was not done by B . Nor would we deny that another person C , the blood donor, is the person who did that past action merely because there is someone other than himself, namely A , who quasi-remembers it from the inside. So here it would not be true that if at time t a person quasi-remembers a past action from the inside then he is identical with the person who did it if anyone existing at t is identical with the person who did it.

Yet even in such a world it seems essential that in any total mental state the memories, i.e., the quasi-memories produced by the past history of the person whose total mental state it is, should outnumber the quasi-memories produced by any given tributary. If the quasi-memories produced by a given tributary outnumbered the memories then surely the tributary would not be a tributary at all, but would instead be the main stream. But this implies that if a person quasi-remembers an action from the inside then, in the absence of evidence to the contrary, he is entitled to regard it as more likely that the action was done by him than that it was done by any other given person. And this, taken together with my earlier point that if someone quasi-remembers an action from the inside there is a presumption that he is the person who did it, gives us a sense in which quasi-memory can be said to provide the quasi-rememberer with "special access" to his own past history. This is of course a much weaker sense of "special access" than that explained in Sect. I—but in this sense it will be true in *any* possible world, and not merely in ours, that people have a special access to their own past histories.

V

In the preceding sections it was assumed that remembering, as opposed to (mere) quasi-remembering, necessarily involves the satisfaction of the strong previous awareness condition; that is, it was assumed that in any genuine case of event memory the memory must correspond to a past cognitive and sensory state of the rememberer himself. And this is commonly supposed in discussions of memory and personal identity. But it is not really clear that this assumption is correct. For consider again the hypothetical case in which a man's body "splits" like an amoeba into two physiologically identical bodies, and in which both offshoots produce memory claims corresponding to

the past life of the original person. Or, to take a case that lies closer to the realm of real possibility, consider the hypothetical case in which a human brain is split, its two hemispheres are transplanted into the newly vacated skulls of different bodies, and both transplant recipients survive, regain consciousness, and begin to make memory claims that correspond to the past history of the brain "donor."²⁰ In neither case can we identify both of the physiological offshoots of a person with the original person, unless we are willing to take the drastic step of giving up Leibniz' Law and the transitivity of identity. But is it clear that it would be wrong to say that each of the offshoots remembers the actions, experiences, etc., of the original person? There is, to be sure, an awkwardness about saying that each offshoot remembers *doing* an action done by the original person, for this seems to imply that an action done by one and only one person was done by each of the two non-identical offshoots. But perhaps we can say that each of the offshoots does remember the action "from the inside." In our world, where such bizarre cases do not occur, the only actions anyone remembers from the inside are those that he himself performed, so it is not surprising that the only idiomatic way of reporting that one remembers an action from the inside is by saying that one remembers doing the action. But this need not prevent us from describing my hypothetical cases by saying that both offshoots do remember the actions of the original person, and it does not seem to me unnatural to describe them in this way. If this is a correct way of describing them, then perhaps my second sort of quasi-remembering, i.e., quasi-remembering, turns out to be just remembering, and the previous awareness condition for remembering turns out to be the causal requirement discussed in the preceding section rather than the stronger condition I have been assuming it to be.

If the suggestion just made about the conditions for remembering is correct, the logical connection between remembering and personal identity is looser than I have been supposing it to be. Yet adopting this suggestion does not prevent one from defending the claim that remembering is constitutive of and criterial for personal identity; on the contrary, this makes it possible to defend the letter of this claim, and not just its spirit, against the very common objection that any attempt to analyze

²⁰ See Wiggins, *op. cit.*, p. 53, where such a case is discussed.

personal identity in terms of memory will turn out to be circular.

Bishop Butler objected against Locke's account of personal identity that "one should really think it self-evident, that consciousness of personal identity presupposes, and therefore cannot constitute, personal identity, any more than knowledge, in any other case, can constitute truth, which it presupposes."²¹ More recently several writers have argued that while "*S* remembers doing *A*" entails "*S* did *A*" (and so entails "*S* is identical with the person who did *A*"), this is only because "*S* remembers doing *A*" is elliptical for "*S* remembers himself doing *A*."²² To offer as a partial analysis of the notion of personal identity, and as a criterion of personal identity, the formula "If *S* remembers (himself) doing action *A*, *S* is the same as the person who did *A*" would be like offering as a partial definition of the word "red," and as a criterion of redness, the formula "If *S* knows that *X* is red, then *X* is red." In both cases the concept allegedly being defined is illicitly employed in the formulation of the defining condition. Likewise, it has been argued that while someone's remembering a past event is a sufficient condition of his being identical with a witness to the event, we cannot use the former as a criterion for the latter, since in order to establish that a person really does remember a given past event we have to establish that he, that very person, was a witness to the event. And if this is so, the formula "If *S* remembers *E*, *S* is identical with someone who witnessed *E*" will be circular if offered as a partial analysis of the concept of personal identity.²³

Such objections assume that remembering involves the satisfaction of the strong previous awareness condition, and they can be avoided on the assumption that the previous awareness condition is weaker than this, e.g., is that given for quasi-remembering in Sect. IV. Or, better, they can be avoided if we explicitly use "remember" in a "weak" sense ("remember_w") rather than in a "strong" sense ("remember_s"), the strength of the

sense depending on the strength of the associated previous awareness condition. Although there are perhaps other possibilities, let us take "remember_w" to be synonymous with "quasi-remember." Clearly, to establish that *S* remembers_w event *E* (or remembers_w action *A* from the inside) it is not necessary to establish that *S* himself witnessed *E* (or did *A*), for it will be enough if *S* is the offshoot of someone who witnessed *E* (did *A*). And while we cannot claim that statements about what events or actions a man remembers_w logically entail statements about his identity and past history, this does not prevent the truth of the former from being criterial evidence for, and from being partially constitutive of, the truth of the latter. For we can still assert as a logical truth that if *S* remembers_w event *E* (or remembers_w action *A* from the inside), and if there has been no branching of M-type causal chains during the relevant stretch of *S*'s history, then *S* is one of the witnesses of *E* (is the person who did *A*). Here we avoid the circularity that Butler and others have thought to be involved in any attempt to give an account of personal identity, and of the criteria of personal identity, in terms of memory.

In the actual world, people remember_s whatever they remember_w, and this makes it difficult to settle the question of whether it is the weak or the strong sense of "remember" that is employed in ordinary discourse. It is possible that this question has no answer; since branching of M-type causal chains does not in fact occur, and is seldom envisaged, people have had no practical motive for distinguishing between the strong and the weak senses of "remember." But I do not think that this question is especially important. We can defend the spirit of the claim that memory is a criterion of personal identity without settling this question, although in order to defend the letter of that claim we must maintain that in its ordinary use "remember" means "remember_w."

At this point I should say something about why it is important to insist on the claim that there is a

²¹ Joseph Butler, "Of Personal Identity," First Dissertation to the *Analogy of Religion*. Reprinted in Flew, ed., *Body, Mind and Death*, (New York, 1964), pp. 166-172.

²² See A. J. Ayer, *The Problem of Knowledge*, (Harmondsworth, Middlesex, 1956), p. 196, and B. A. O. Williams, "Personal Identity and Individuation," in Gustavsen (ed.) *Essays in Philosophical Psychology* (New York, 1964), pp. 327-328 (originally published in the *Proceedings of the Aristotelian Society*, vol. 57, 1956-57).

²³ See Williams, *op. cit.*, p. 329, and my *Personal Identity and Memory*, pp. 869-870 and p. 877. In the latter, and in *Self-Knowledge and Self-Identity*, I attempt to reduce the force of this objection by arguing that it is a "conceptual truth" that memory claims are generally true, and that we can therefore be entitled to say that a person remembers a past event without already having established, or having inductive evidence, that some other criterion of personal identity (one not involving memory) is satisfied. This way of handling the objection no longer seems to me satisfactory.

causal element in the notion of memory. For this claim has recently come under attack.²⁴ It has been argued that the notion of memory should be analyzed in terms of the *retention*, rather than the causation, of knowledge, and that the notion of retention is not itself a causal notion. Now I have no objection to saying that remembering_s consists in the retention of knowledge. But I believe that unless we understand the notion of retention, as well as that of memory, as involving a causal component, we cannot account for the role played by the notion of memory, or even the concept of similarity, in judgments of personal identity.

Here it will be useful to consider a hypothetical case I have discussed at some length elsewhere.²⁵ Let us suppose that the brain from the body of one man, Brown, is transplanted into the body of another man, Robinson, and that the resulting creature—I call him “Brownson”—survives and upon regaining consciousness begins making memory claims corresponding to the past history of Brown rather than that of Robinson. We can also suppose that Brownson manifests personality traits strikingly like those previously manifested by Brown and quite unlike those manifested by Robinson. Although Brownson has Robinson’s (former) body, I doubt if anyone would want to say that Brownson is Robinson, and I think that most people would want to say that Brownson is (is the same person as) Brown.

But what can we offer as evidence that Brownson is Brown? Clearly the mere correspondence of Brownson’s ostensible memories to Brown’s past history, and the similarity of Brownson’s personality to Brown’s, is far from being sufficient evidence. And it is equally clear that the notion of the *retention* of knowledge and traits is of no use here. To be sure, once we take ourselves to have established that Brownson is Brown we can say that Brownson retains knowledge, and also personality traits, acquired by Brownson in the past. But the latter assertion presupposes the identity of Brownson and Brown, and cannot without circularity be offered as evidence for it. Indeed, the circularity is the same as what would be involved in offering as evidence of this identity the fact that Brown-

son remembers_s Brown’s past experiences and actions.

We do not, however, beg the question about identity if we take Brownson’s possession of what used to be Brown’s brain, together with the empirical facts about the role played by the brain in memory, as establishing that Brown’s ostensible memories are directly M-connected with Brown’s past actions and experiences, i.e., are causally related to them in essentially the same ways as people’s memories are generally connected with their own past experiences and actions. This in turn establishes that Brownson quasi_o-remembers, and so remembers_w, Brown’s past experiences and actions. And from this in turn, and from the fact that we have good reason to suppose that no other person’s memories are M-connected with Brown’s past history in this way, i.e., that there has been no “branching” of M-type causal chains, we can conclude that Brownson is Brown.²⁶

We can reason in this way only if we can assert that there is a causal connection between Brownson’s past history and Brownson’s ostensible memories. And this, it seems to me, we are clearly entitled to do. Given that Brownson has Brown’s former brain, there is every reason to think that had Brown’s history been different in certain ways, there would (*ceteris paribus*) be corresponding differences in what Brownson ostensibly remembers. I can see no reason for doubting that such counterfactuals assert causal connections. Similar remarks can be made about the similarity between Brownson’s and Brown’s personality traits. Given that Brownson has Brown’s former brain, we have reason to think that had Brown developed a different set of personality traits, Brownson* would (*ceteris paribus*) have those personality traits rather than the ones he has. And while we cannot naturally speak of Brown’s having a certain trait at one time as causing Brownson to have the same trait at a subsequent time, we can speak of the former as being an important part of a causally sufficient condition for the latter. It is only where we suppose that the traits of things at different times are causally related in this way that we are entitled to take the similarity of something at one

²⁴ See Squires’ “Memory Unchained,” *op. cit.*

²⁵ *Self-Knowledge and Self-Identity*, pp. 23–25 and 245–47.

²⁶ In *Self-Knowledge and Self-Identity* I held that saying that Brownson is Brown would involve making a “decision” about the relative weights to be assigned to different criteria of personal identity, and that in the absence of such a decision there is no right answer to the question whether Brownson is Brown. I have come to believe that there is a right answer to this question, namely that Brownson is Brown, and that my former view overlooked the importance of the causal component in the notion of memory—see my treatment of this example in “On Knowing Who One Is,” *op. cit.*

time and something at another time as evidence of identity.

VI

We are now in a position to reassess the view, mentioned in Sect. II, that the knowledge of our own pasts and our own identities provided us by memory is essentially "noncriterial." If I remember_s an action or experience from the inside, and know that I do, it makes no sense for me to inquire whether that action or experience was my own. But it seems logically possible that one should remember_w an action or experience from the inside (i.e., quasi_c-remember it) without remembering_s it. So if one remembers_w an action or experience from the inside it can make sense to inquire whether it was one's own (whether one remembers_s it), and it would seem offhand that there is no reason why one should not attempt to answer this question on the basis of criteria of personal identity.

But while an action I remember_w from the inside can fail to be mine, there is only one way in which this can happen, namely through there having been branching in the M-type causal chain linking it with my present memory. So in asking whether the action was mine, the only question I can significantly be asking is whether there was such branching. If I go on to verify that there was no branching, I thereby establish that a sufficient criterion of personal identity is satisfied. If instead I conclude on inductive grounds that there was no branching, relying on my general knowledge that M-type causal chains seldom or never "branch (or that it is physiologically impossible for them to do so), I thereby conclude that a sufficient criterion of personal identity is satisfied. But an important part of what the satisfaction of this criterion consists in, namely my remembering_w the past action from the inside, is not something I establish, and not something I conclude on inductive grounds, but is something I necessarily presuppose in inquiring concerning my relation to the remembered_w action. In cases where one remembers_w a past action from the inside, and knows of it only on that basis, one cannot significantly inquire concerning it whether one does remember_w it—for as I tried to bring out in my discussion of quasi-remembering, there is no way of knowing the past that stands to remembering_w as remembering_w stands to remembering_s, i.e., is such that one can know of a past event in this way and regard it

as an open question whether in so knowing of it one is remembering_w it. So in such cases the satisfaction of this part of the memory criterion for personal identity is a precondition of one's being able to raise the question of identity, and cannot be something one establishes in attempting to answer that question.

That one remembers_w a past action is not (and could not be) one of the things one remembers_w about it, and neither is the fact that there is no branching in the M-type causal chain linking it with one's memory of it. And normally there is no set of remembered_w features of an action one remembers_w from the inside, or of the person who did the action, by which one identifies the action as one's own and the agent as oneself. If one has not identified a remembered person as oneself on the basis of his remembered_w features, then of course it cannot be the case that one has misidentified him on this basis. This is not to say that there is no basis on which one might misidentify a remembered_w person as oneself. If there can, logically, be remembering_w that is not remembering_s, then where one remembers_w an action from the inside one's judgment that one did the action will not be logically immune to error through misidentification in the sense defined in Sect. II—though given the contingent fact that all remembering_w is remembering_s, such judgments can be said to have a *de facto* immunity to error through misidentification. But the sort of error through misidentification to which a statement like "I saw a canary" is liable, if based on a memory_w from the inside, is utterly different from that to which a statement like "John saw a canary" is liable when based on a memory_w of the incident reported. If the making of the latter statement involves an error through misidentification, this will be because either (1) the speaker misidentified someone as John at the time the reported incident occurred, and retained this misidentification in memory, or (2) at some subsequent time, perhaps at the time of speaking, the speaker misidentified a remembered_w person as John on the basis of his remembered_w features. But if I remember_w from the inside someone seeing a canary, and am mistaken in thinking that person to have been myself, it is absurd to suppose that this mistake originated at the time at which the remembered_w seeing occurred. Nor, as I have said, will this be a misidentification based on the remembered_w features of the person who saw the canary. What could be the basis for a misidentification in this case is the mistaken belief that

there is no branching in the M-type causal chain linking one's memory with the past incident. But a misidentification on this basis, while logically possible, would be radically unlike the misidentifications that actually occur in the making of third person reports.

VII

Because I have taken seriously the possibility of worlds in which M-type causal chains sometimes branch, and thus the possibility of quasi-remembering (remembering_w) that is not remembering_s, I have had to qualify and weaken my initial claims about the "special access" people have to their own past histories. But if our concern is with the elucidation of our present concept of personal identity, and with personal identity as something that has a special sort of importance for us, then it is not clear that the possibility of such worlds, and the qualifications this requires, should be taken as seriously as I have taken them. For there is reason to think (1) that some of our concepts, perhaps including the concept of a person, would necessarily undergo significant modification in their application to such worlds, and (2) that in such worlds personal identity would not *matter* to people in quite the way it does in the actual world.

There are important connections between the concept of personal identity and the concepts of various "backward looking" and "forward looking" mental states. Thus the appropriate objects of remorse, and of a central sort of pride, are past actions done by the very person who is remorseful or proud, and the appropriate objects of fear and dread, and of delighted anticipation, are events which the subject of these emotions envisages as happening to himself. And intentions have as their "intentional objects" actions to be done by the very person who has the intention. It is difficult to see how the notion of a person could be applied, *with these conceptual connections remaining intact*, to a world in which M-type causal chains frequently branch, e.g., one in which persons frequently undergo fission. If I remember_w from the inside a cruel or deceitful action, am I to be relieved of all tendency to feel remorse if I discover that because

of fission someone else remembers_w it too? May I not feel proud of an action I remember_w from the inside even though I know that I am only one of several offshoots of the person who did it, and so cannot claim to be identical with him? Am I not to be afraid of horrible things I expect to happen to my future offshoots, and not to view with pleasant anticipation the delights that are in prospect for them? And is it to be impossible, or logically inappropriate, for me knowingly to form intentions, and make decisions and plans, which because of the prospect of immanent fission will have to be carried out by my offshoots rather than by me? To the extent that I can imagine such a world, I find it incredible to suppose that these questions must be answered in the affirmative. The prospect of immanent fission might not be appealing, but it seems highly implausible to suppose that the only rational attitude toward it would be that appropriate to the prospect of immanent death (for fission, unlike death, would be something "lived through"). It seems equally implausible to suppose that a person's concern for the well-being of his offshoots should be construed as altruism; surely this concern would, or at any rate could, be just like the self-interested concern each of us has for his own future well-being. Yet a negative answer to my rhetorical questions would suggest that either the concept of a person or such concepts as those of pride, remorse, fear, etc., would undergo significant modification in being applied to such a world.²⁷

A person's past history is the most important source of his knowledge of the world, but it is also an important source of his knowledge, and his conception, of himself; a person's "self-image," his conception of his own character, values, and potentialities, is determined in a considerable degree by the way in which he views his own past actions. And a person's future history is the primary focus of his desires, hopes, and fears.²⁸ If these remarks do not express truths about the concept of personal identity, they at least express truths about the *importance* of this concept in our conceptual scheme, or in our "form of life." It seems plausible to suppose that in a world in which fission was common personal identity would not have this sort of

²⁷ On this and related questions, see my exchange with Chisholm in *Perception and Personal Identity*, *op. cit.*

²⁸ This is not to deny the possibility or occurrence of unselfish attitudes and emotions. Even the most unselfish man, who is willing to suffer that others may prosper, does not and cannot regard the pleasures and pains that are in prospect for him in the same light as he regards those that are in prospect for others. He may submit to torture, but he would hardly be human if he could regularly view his own future sufferings with the same detachment (which is not indifference) as he views the future sufferings of others.

importance. Roughly speaking, the portion of past history that would matter to a person in this special way would be that which it is possible for him to remember_w, and not merely that which it is possible for him to remember_s. And the focus of people's "self-interested" attitudes and emotions would be the future histories of their offshoots, and of their offshoots' offshoots, and so on, as well as

their own future histories. In the actual world it is true both that (1) remembering_w is always remembering_s (and thus that there is special access in the strong sense characterized in Sect. I), and that (2) the primary focus of a person's "self-interested" attitudes and emotions is his own past and future history. It is surely no accident that (1) and (2) go together.²⁹

The Rockefeller University

Received December 9, 1969

²⁹ This is a considerably revised version of a paper which was read at a conference on "The Concept of a Person" at the University of Michigan in November 1967, and at the University of British Columbia and the University of Saskatchewan at Saskatoon in the Spring of 1969. I am grateful to Harry Frankfurt, Robert Nozick, and Michael Slote for criticisms of the earlier versions of the paper.

II. THE THEORY THAT THE WORLD EXISTS BECAUSE IT SHOULD

JOHN LESLIE

I. EXPLAINING ALL EXISTENTS

IT is widely agreed that one cannot explain why there exists something rather than nothing. Everything explicable must, it is said, be explained through something beyond; hence to explain everything we must propose an infinite chain, each link created by its predecessor. But does not this propose that explanation be deferred infinitely? Second, how is each link bound to the next? The quick answer, that it "necessitates" it, is too vague. Third, to suggest an endless regress of necessitating things is boundlessly obstinate.

Well, yes, one cannot explain everything; but might not one account for every existing object? Suppose one explained the world through this alleged reality: that it ought to exist. Would an ethical right to exist be a separate existent? It is better called a status which existents perhaps have. The sum of all existing objects might owe its presence to the ethical reason for existence which such a status provides.

Axiarchism is my label for all theories picturing the world as ruled by value. One theory stands out: that the world's existence and detailed nature are products of a directly active ethical necessity. This rejects reliance on the inexplicable creative prowess of an inexplicably existing benevolent deity. However, it permits belief in God, who may himself be ethically required; or (with some theologians) "God" may be the name of the principle that ethical requirements are creatively powerful. I shall defend axiarchism of this extreme variety: the theory that the sum of things has an ethical requiredness which is—as a matter not of logic but of fact—enough to guarantee its existence.

II. MORAL STRUGGLE IN AN OPTIMIST UNIVERSE

Extreme axiarchism must be optimism, the faith that the world is the best possible. But surely optimism presents goodness in embarrassing quan-

tity. If the world is as good as can be, what room is left for moral effort?

One reply appeals to absolute freedom of the will. It maintains that choices are irreducibly open; not matters of mere chance, yet undetermined; which is a supreme good, to be expected in an ethically ideal universe. This allows for morality by placing strong restrictions on the sense in which the world is the best possible. Thus if we were to put back all clocks, the course of events, obeying absolute freedom's dictates, would be unlikely to repeat itself; it might run worse, but would remain better than any possible tram-line process. However, the concept of freedom presented here may be confused, and optimist defense of moral sweat need not recruit it.

We can think both that the course of events is guaranteed the best logically possible and that we have moral duties. This seems excessively paradoxical. Yet for moral aims to have a place, what is essential is that we should be *in some sense* "free" to push events down any of several paths. Now, for events to be guaranteed best logically possible, what is essential is *some sense* in which we are not "free." Many philosophers accept that all events, including our decisions, are determined by scientific laws, which perhaps provides some ground for dubbing ourselves "machines" and "not free." But that gives no cause for the fatalist doctrine that the world pursues the same course whether we will or not, neither does it give clear cause for describing our will as thrust upon us. If machines, we are machines for complex deliberation, which surely supplies some excuse for talk of "freedom."

Our brains may perform thought processes when their atoms are governed by deterministic natural forces (I mean, forces deterministic at the macroscopic level of thought processes: whether they are statistical at some microscopic level is irrelevant). But if so, *we* are not tyrannized by those forces, for the forces, just as much as the atoms, then enter into our mental make-up; if they were to change radically, brains might remain, but we should dis-

appear. Being essential to our mental identity—constituents of our minds—the forces governing our brains could not also stand outside our minds to tyrannize them, for that would be to govern our decisions twice over. Again, that our decisions are thus governed by natural forces does not deny that they are governed by thought processes: see the start of this paragraph. Now, those processes ought to include moral deliberation. This is all repeatedly climbed rock, but less well recognized is its relation to optimistic castles. If the optimist believes—and why not?—that the best logically possible world is scientifically deterministic, he can find use for moral effort, as can any scientist.

I am not saying that someone in the best logically possible pattern could make that pattern better than best possible. Yet his decisions, as parts of that pattern, causally connected with other parts, would affect its goodness. I mean: the decisions which a person with his mental make-up would make are factors correlated with just how good (that is, how much better than any individual thing) “best logically possible” is. Decisions, or wishes, or potatoes, if part of a best possible world, would affect its goodness from inside, without making it better or worse than best; just as the number seven affects (is correlated with) the sum of numbers one to ten, without making that sum greater than the sum of those numbers. Certainly the character and value of events in the best logically available plan is a matter determined from the dawn of creation, but the same could be said of the character and value of decisions made by minds wholly governed by deterministic laws.

What scene should we expect, if optimism were right? Concentrating on earthquakes and famines, we protest at the deity or principle of value allegedly responsible—and forget the only very clear alternative, the trivial arbitrariness of a drug addict’s dream. One can fight disease and fire and unpleasant people while welcoming that their absence is not magically ensured, as life’s interest is in its working without magic. (Perhaps there could be a universe better than ours which yet obeyed scientific laws throughout; there is no obvious contradiction here; but is there contradiction in a sixteenth prime number between one and fifty-one? The question involves no dubious value judgments, and does not ask us to survey universes. Half-minute obviousness?)

What must be remembered is how agonizingly bad our best logically possible world might be: not worse than nothing, but bad compared with

what we might have made of it. The interest of scientific order, the drama of free action, may be supreme goods, but are had at an often terrible price—a price we can reduce, for the determinateness of a best logically possible system need not involve a sterile fatalism.

Such defense against attack from the side of morality is not open to all optimists. It maintains that one great ethical requirement can, unfortunately, conflict with another still greater, and be overruled. In our best possible world, the greater always determines realities, but it remains up to us to act so that other requirements are satisfied as well. For a supreme requirement is that we should have the freedom allowed in a world scientifically ordered throughout; or perhaps some more mysterious freedom. We must ensure, *by freely acting well*, that the requirement that we should have freedom does not conflict with, and so cannot overrule, the requirement that we should act well. But this notion of *overruling* is disallowed by many optimists. For instance, the Hegelian philosophy perfected by Bradley makes the world, not a best possible compromise, but absolutely flawless, a world in which evil and conflict of goods are appearance only. That seemingly renders effort pointless.

III. VALUE AS MAPPABLE FACT

The extreme axiarchist, who sees ethical requirements as themselves creative, must, when he describes things as “in fact good,” picture himself as *mapping* realities, and not, for example, as merely issuing prescriptions inciting us to various actions toward realities.

My objection to prescriptivism (and emotivism and similar theories) concerns the difficulties which its supporters face in defending any complete set of ethical standards. Any single standard they can defend as integrating well with others, but when a complete set is at stake this defense becomes that the standards prop one another up. That is a poor substitute for belief that in describing certain standards as good we correctly map facts of goodness. Admittedly, a prescriptivist can speak of a “descriptive element” in his prescriptions, yet this means only that his standards impose fixed preferences for various mappable characteristics when he takes some scientific or other map of the world and marks certain areas “good or prescribed.” His “descriptions” of goodness thus take into account mappable facts, but do not themselves map them.

Hence no prescriptivist can forcefully support his set of standards with the claim that it guides toward things themselves good; for, though he can make that claim, he views it only as a prescription, based either on nothing or *on the standards themselves*, of the things which the standards favor.

In contrast, forceful support is provided by the hypothesis that certain standards guide toward mappable realities of goodness. For this enables us to refer moral rules to something beyond themselves, to which they are responsible, much as rules for grading apples are responsible to market demands. We may be unembarrassed to admit that this "something beyond" is realities not readily verifiable: compare unembarrassed support for church-going by reference to unverifiable maps of heaven and hell.

Yet suppose—the prescriptivist may counter—that faith in mappable ethical realities came to be lost. Could the loss give point (worth) to sitting depressed, even granted that one's new outlook made all things pointless (worthless) from the old viewpoint? That moral enthusiasm *ought* to find justification in realities beyond the enthusiasm itself is an assumption I am compelled to accept only if already believing that those extra realities are the source of all *oughts*. Remember that we are logically necessitated to follow some policies for action, as even sitting depressed is acting somehow. Given this necessity to favor some policies or other, what further excuse need we demand?—However, such a counter attempts to be too strong. It gives a logical reason for having preferences: hence any agent, in any logically possible world, would have that reason. Yet, one can protest, some of those logically possible worlds would contain nothing good or bad, or would be ones where every choice was precisely as good or bad as every other. When, in such ethical grayness, the prescriptivist would retain his excuse for rejecting "further support" for moral enthusiasm, perhaps *that* is why we should search for such support in mappable realities.

There are, however, strong arguments against such search. Before tackling these, some misunderstandings require removal. First, to call value intrinsic to things as a matter of mappable fact does not deny that likes and dislikes are ethically significant, for it might be that the sole objects having intrinsic value were states of enjoyment. Second, recognition of intrinsic goodness by no means ends moral discussion, for the need for an intrinsically good thing—a thing in itself better than nothing—

can be overruled by the need to achieve a more intrinsically good alternative or to avoid intrinsically undesirable consequences.

Next, more serious difficulties. One drawback with such primitive theories as that "good" means "pleasant" is that the idea of an existent property seems quite different from that of an ethical claim to exist. Demonstrably, questioning whether pleasure is good is not questioning whether pleasure is pleasurable. Admittedly some modern "ethical naturalists" have reconsidered such demonstrations. They argue that, though "good" does not mean, simply, "pleasant," pleasure is strongly relevant to talk of goodness; that pleasure, though at times to be sacrificed, is one of many ingredients of goodness; and that anyone disagreeing has missed how the word "good" is used. Yet what appears to me most characteristic of our use of "good" is that goodness is understood to be sufficient grounds for action: hence what seems suggested by these modern naturalists is that "sufficient grounds for action" means, simply, "promotion of pleasure, peace, health, etc." But this particular account of how language is used does not by itself even begin to inspire us to do anything. (It is an odd conversation which runs: "What is it good to do?" "*Why, things like doing as you would be done by; spreading joy; some compromise between the things normally called good.*" "But why good to act so, instead of hunting for haddock's eyes?" "*That's just how the word 'good' is used. You may not want to be 'good,' making it linguistically correct to call you 'bad.'*") . . . Now, much the same drawback seemingly confronts absolutely all theories which make value a mappable reality.

Thus, it can seem no advance to replace natural properties such as pleasurable (or those medleys of properties favored by modern naturalism) by a less readily detectable "quality of goodness" revealed by "intuition," or even by an undetectable property postulated by faith. The distinction between properties of existents and ethical claims to exist still plagues us.

Another problem is peculiar to theories which make goodness not readily verifiable. We favor situations for their visible qualities, looking on features like pleasurable as good themselves, not as signs of something hidden and alone the proper object of ethical awe, as redness is a sign of heat in iron.

Attempts to counter these problems involve speaking of goodness as a property derivative or consequential, following from more ordinary pro-

perties with absolute necessity. All too obscure. What is the unbreakable ("absolutely necessary") bond? Suppose it is said that goodness is not a "constitutive" property like redness which makes things what they are. One asks: How else does it achieve reality, and does it show how goodness is bound to other properties if one specifies what this special property *is not*?

A stonewall answer could be that, though goodness can be called a property, it is better named a property-of-properties or ethical status possessed through possessing various properties. That, however, has something of the air of a purely verbal escape. I shall try a further approach, suggesting an analogy between goodness and certain relations.

IV. THE ETHICAL BEARING OF NATURE ON EXISTENCE

Relations occupy two main divisions. One includes: standing on the same mushroom; annoying the same politician; being separated by an elephant. In contrast are: being twice as long as; having a color nearer to red than; being the same shape as. Though relations in the first division could be thought able to change while the related entities remained unchanged, the same is not true of those in the second. Thus two pebbles, standing on a mushroom, perhaps remain just the same if fallen from their perch. But when two things are related by similarity of shape, they could not fall out of this relation if both remained unaltered. Again, when an orange splodge is nearer to being a red than a yellow splodge, the three could not change order if remaining unaltered splodges.

An explanation is easily found. When a relation between entities can vary without their variation, some further, completely separate reality is involved: mushroom, politician, elephant. (Varying relations of spatial separation may seem exceptions, but there are ways of denying this: for example, space itself may be a separate something, as talk of matter as "wrinkles in space-time" suggests.) In the opposite case, no completely separate reality is present. When two realities are similar in shape, no doubt the similarity might be classified as another reality, but there is nothing truly separate here. Again, when one object is twice as long as another, this being twice as long is hardly a third object, nor does it involve any third object. Relations like these last are *wholly secondary in their reality to the reality of the related terms*. (Somewhat as an object's shape is secondary to that object: an

aspect of it) but not, like shape, a constitutive aspect.) Their existence is entirely derivative and consequential, and hence follows from that of the related entities with absolute necessity. All this is familiar and now not mysterious. Even "absolute necessity" becomes easy enough. A relation deriving its reality solely from the related entities could never fail to connect them, if they remained unaltered; for a relation able to come and go as it pleased would not be one whose reality was wholly secondary to the related things, but would be rather like a separately existing chain coming and going between unaltered trucks. (Not quite like a chain, for it could never survive in the absence of the things connected: still, its reality would include some element unambiguously additional to theirs.)

Yet what relevance has this to mappable intrinsic value? Which things should be mapped as in relation to one another? Well, a thing of intrinsic value can be viewed as related to itself (compare how the number one is its own square). It is self-justifying, in contrast to things justified by their effects, which says that it stands to itself in the ethical relation "providing some degree of justification for." (As there are degrees of good, so there are degrees of justification. The relation comes in various strengths, as do relations of similarity, proximity, and so on.)

Further detail is suggested when we call things "such as ought to exist." We might picture the *nature* of any intrinsically good thing as standing in the relation of justifying that thing's *existence*. . . . An attempt to revive some unhelpful mythology? Existence the ermine of the privileged in a realm of essences? Well, nothing much is intended beyond the initial position: that intrinsically good things stand related to themselves. To say that a thing's nature provides justification for its existence invites us to consider the thing in two ways. First, we ask what it is; we consider it with respect to "its constitution." And next, whether there ought to be a thing of that sort; which considers it with respect to "its existence." The relation is, therefore, between the thing (*qua* possessing a certain constitution) and itself (*qua* existent object). (Note that some realities are *not* existent objects: e.g., the real impossibility of trisecting most angles with ruler and compass, and the real possibility of bisecting them.)

When such a relation ("quasi-relation," perhaps) is identified with value, that explains more than why value is consequential or derivative, following from other properties with absolute

necessity. It explains also why it is not a constitutive property, for the relation is between a thing, as having the constitution which it has, and itself.

The relation would be invisible, since our senses detect how things are constituted, not how they stand to themselves ethically through being so constituted. Mappable intrinsic value is a matter for belief. We can argue persuasively about it, tracing interrelations of various hypotheses, but can never prove its reality.

V. BEING MARKED OUT FOR EXISTENCE

The need for belief may be found unalarming (as with religion), yet the invisibility of value does bequeath a serious problem—of meaning. How can words come to life, if not through confrontation with sensations?

In answer, I rely on a limited analogy between ethical requirements and causally effective requirements. At times we have no say as to what shall exist, yet at others we can choose; now, the believer in mappable value, though welcoming choice, is in a way uneasy at it. He is unhappy with justifications of an act which merely point out that it reflects well-integrated preferences, or that most men would name it "good." Suppose next (which simplifies matters) that he considers some act as justified by its intrinsic nature, rather than by its effects. He pictures it as marked out for existence by that nature, somewhat as various things are marked out by realities which effectively compel, for he wishes to serve a selecting factor beyond his own or other people's moral tastes. He looks to realities beyond those which effectively force our hands, but which somewhat similarly limit the field, so that, while many acts are on the cards as practical possibilities, only a few are "really on the cards"; some things one "has to do," others one "just cannot do." Though in one sense he is at liberty to do bad—though he certainly will do bad on occasion—he looks for a sense in which he lacks that liberty. There obviously is an aspect in which any act effectively possible is possible whatever ethics has to say, but he wishes there to be some way in which his hands can be described as nonetheless forced.

Now such descriptions lose their paradoxical air when it is seen that moral freedom to act just as he pleases is what he wishes to lack. However, he can add depth to talk of moral needs by describing his situation paradoxically. *He can show what he means by ethical factors, by saying they force his hands somewhat as (though not precisely as) do effective factors—and not*

that they "force his hands ethically," which provides only a circular characterization.

All this has nothing to do with the two senses of words like "free" and "forced" which a determinist might accept when discussing free will. Let me stress that ethical necessities might be always and entirely without effective influence, i.e., I understand "necessity" in a way permitting this. By all means replace "having an existence ethically necessary" by "being ethically designated for existence." All that is important is that there is some *analogy* between being effectively forced to do something (by neurotic compulsion, say) and being ethically forced: an analogy not holding between being effectively forced to do something and having any sort of taste for it.

The analogy must be defended against attempts to make ethical requirements somehow hypothetical. Someone may urge that there is no ethical necessity more dramatic than this: that if we are going to behave morally, then *in view of that it is necessary* to act in some specified fashion in any given situation. But that makes ethical demands too relative. Compare the parody that, if we are to make the world red, then it is necessary to buy vast quantities of paint; or that if we are to produce an ugly world, then it is necessary to introduce massacres.

Philosophical maneuvers to give meaning through analogy are suspect, since the analogies are typically so partial. However, sufficient basis for meaning is, I feel, gained by describing ethical requirement as analogous to effective requirement in that each narrows the field open to us non-hypothetically, though only one is defined as narrowing it with practical success, and the precise way in which the other must narrow it is unique. Ethics is a difficult region, and it would not be too surprising if its basic concept barely squeezed into the sunlight of the meaningful, where "prescribed" and "named good by the majority" bask so confidently. The difficulty is in distinguishing an act's having certain verifiable features (of pleasurable-ness, or of spreading redness, or whatever) from its being marked out for existence, and the analogy between ethical and effective necessities is essential for giving meaning to the distinction, since such phrases as "marked out for existence" are no magnets to draw meaning to themselves.

VI. NEEDS TRANSCENDING NEEDS FOR ACTION

What about ethical needs which are not demands that someone act in a specified way? (Or is that

nonsense? It would certainly seem nonsense to a prescriptivist, or given "duty" as the basic ethical concept.) This issue is central to an extreme axiarchism. In the absence of all objects, and hence of all agents, what would be the status of the need for a good world to exist? Would it be nothing? Or might ethical hypotheticals—facts about what would be good—possess importance when there were no gods or persons whatever? Could they possess even creative importance?

We might effect some divorce between ethical necessity and the analogy used to introduce it, an analogy concerning the situation where an agent wants reasons for action. Might there not be recognizably ethical necessities outside such situations? While we view some acts as justified by their intrinsic nature alone, others we see as justified mainly or entirely by their effects. We often feel that, if it is necessary to favor, say, a thing's continued existence, then that is because the thing is already necessary: already marked out for existence in a way going beyond the goodness of acts favoring it, and itself the source of their goodness. This pictures a thing's ethical necessity as spreading to acts which are its causal prerequisites like dye across damp cloth. And the picture is perhaps not so unusual, for ethics is not the sole area of necessities able to spread: causation and logic provide long chains, the necessity of each link yielding that of the next.

Agreed that, if a situation is ethically required, this means required in a manner which would make necessary various acts, if they were its prerequisites. But that comes nowhere near admission that all ethical necessities are really necessities of action; and there do look to be matters ethically important besides that of how people act. If there existed no living things, the materialization of a good world of people would itself be a good development, an answer to an ethical call, rather than a mere prelude to good developments. It can therefore appear reasonable to see value as a status of being-marked-out-for-existence possessed by certain things (e.g., the world) regardless of whether any outside person can favor them. Their nature stands to their existence in the direct relation of justifying it or making it ethically necessary; no person from beyond appears in the reality of the relation. More simply, they are self-justifying, or ethically self-necessitating.

VII. THE LEAP TO AN EXTREME AXIARCHISM

One can now move to axiarchism of a thorough-

going variety. It is entirely speculative: if it corresponds to reality, then that is unverifiable fact, containing an inexplicable element. The importance of ethical necessities may transcend not only the existence of people able to act, but also that of everything else. In an emptiness of all normally describable as existents or objects, there could be ethical realities. For instance, the reality that this emptiness was a tragic lack of something for which there was a need: a best possible world. Axiarchism may view such a reality as creative—relying on the parallel between ethical and effective marked-out-ness to give some charm to this, despite utter failure to guarantee its correctness.

The axiarchist can stress the unconditional ethical significance of possibilities. A mere possibility is not an object (in any normal sense), but that is no cause for disregarding it. Ethics considers, not just what there is, but what might be put in its place. Naturally, common-day ethical disputes do not touch whether what-might-be-so would retain importance if there existed no objects whatever. But suppose an annihilation machine, able to blot out all, itself included, with its inventor justifying its use as follows. (a) Starting the machine will be a pleasure. (b) The consequent absence of objects will be nothing unfortunate; for, if unfortunate, then important, and nothing can have importance when no objects exist. The enormity of his conclusion—that it would not matter that the annihilation button had been pressed—is sufficient refutation. An utter lack of objects need not itself be an object, to be tragic.

No convincing axiarchism would propose that there ever had been such utter emptiness, since if ethical requirements ever were ineffective in creating, presumably they would have continued so. Yet an extreme axiarchism must pronounce on the imaginary situation in which no objects exist, for a theory explaining the existence of all objects must indicate a necessity able (if called on to do so) to create from nothingness—a necessity which axiarchism finds in the unconditional ethical need for a world of a certain type. But what is this need, if not a real object? Well, it is odd to maintain that the absence of good objects would be unimportant in the absence of all objects, and that provides the answer. The unconditional reality of ethical necessity just is the unconditional importance of ethical hypotheticals. In describing these as "important" we no more describe existing objects than in describing them as "without importance," since

hypothetical facts are not in any normal sense objects.

For it to be *real* that a world of a certain sort would be good, it is enough that the constitution of such a world *would indeed* make it good (goodness being tied to other properties as discussed). There is no call for any person or thing to record, meditate on, or otherwise add body to this hypothetical. It could be a reality without being an existent object or concerned with actually existent objects, as it is sufficiently "to do with existents" to be real if it concerns what might exist. When an axiarchist explains all objects through the eternal importance of such a reality, he will not be scared off with the news that realities timeless and bodiless are unfashionable.

One can now identify the creator. In the absence of all objects, the factor necessitating the materialization of a best possible world is the constitution which such a world would have. What is such a constitution but a hypothetical non-existent abstraction? Well, axiarchism may not share prejudice against the merely hypothetical. It is ethically important, so why not creatively so?

But perhaps this reference to a "constitution" screens nonsense as here no firm wedge can divide an object from "its constitution." That a thing's constitution justifies or makes ethically necessary the thing's existence means that the thing is self-justifying—or self-necessitating. And can we take seriously the tale that the world creatively necessitates itself? Next we shall lend credence to Münchhausen's self-levitation by tugging on his pigtail. . . . No, for the axiarchist can do what Münchhausen cannot, citing the unconditional significance of a hypothetical fact: that a good world would have in itself an ethical reason for its existence. Axiarchism may recognize no call to explain every object by another existing previously to it, for why should necessitating factors exist beyond and before the things they necessitate? Ethical (and possibly creatively effective) necessity is not an affair of events-of-type-P-regularly-preceding-events-of-type-Q, which is what a scientist may mean by necessity. The *necessity* for a good world would be "prior"—in time, or at least in principle—to any such world, as it transcends the existence of everything; yet this does not say that the *necessitating factor*—the "source" of the necessity, the ethically desirable constitution which such a world would possess—could exist prior to such a world. (To suggest that ethical *necessities* would be real enough, even in a blank, to carry

creative weight, is just to recognize the importance, even in a blank, of facts about what constitutions would make necessary the existence of any objects which possessed them.)

In short, asking why any objects exist is not a self-contradictory quest for a creative object beyond all objects. An unconditional call for a world of a certain sort may originate in the nature which such a world would have—or has; and that call may be creatively sufficient.

VIII. THE INEXPLICABLE ELEMENT

But why respect such idle speculation?

Well, that the world is *not* due to the creative sufficiency of ethical requirements is quite as idly speculative. Uncomfortably, every creation story must contain features to be taken on trust. One can make the entire cosmos such a feature, theorizing that there is no cause for its presence. Alternatively, ethical requirements may provide a cause; but their alleged effectiveness must be in part inexplicable.

What advantages has the axiarchist story over, say, the strange tale that things exist because they have mass? Well, when a world has value, its continued existence is "marked out" inside the field of logical possibilities, whereas nothing equivalent follows from a world's being massive. Again, the ethical indicatedness which would be had by a world of a certain sort is an affair of unconditional and direct importance; whereas that some world would be massive has, at most, an importance derived from some devious tie between massiveness and goodness. If one is in the mood, all this can add charm to the alleged effectiveness of value; yet plainly there is great room for disagreement over how much charm it does add. So I can (just about) sympathize with those who—though accepting that the universe *may* be best possible—nonetheless judge it not the least more likely that it exists because ethically required, than that it exists because containing massive objects, or unwavering sceptics, or because it is an ethical disaster.

No, I am not proclaiming that axiarchist reasoning lacks force; only that rationality does not compel acceptance of its force. Properly developed, it is reasoning, not a pun on "necessity," but reasoning whose principles come with the apology: these persuade me, and might persuade you. We have two rightly separable concepts, of things ethically marked out for existence and of things effectively

marked out. Nonetheless, the extreme axiarchist can reasonably speculate that, in the reality which these concepts mirror, a relation might be one of *marking out for existence ethically and with creative effectiveness*.

Such a relation supposedly connects the nature and existence of the world in its supremely desirable entirety. It is a single, fully unified relation; its creative aspect manages to be real only because it is an aspect of the whole which it makes when combined with the ethical aspect; for the theory is, not just that the world exists *and* this is best, but that it exists *because* this is best. (Though not "simply because" of that, for the creative aspect remains distinguishable from the ethical one, and "simply" might seem to deny this. "Because" indicates that *ethical requiredness* is a prerequisite of *creative ethical requiredness*.) Bitter experience shows how hard it is to induce professionals to distinguish this from the ludicrous positions (a) that logic guarantees that ethical necessities are creatively active, and (b) that the world simply happens to be best possible. There is a sense in which ethical reasons for existence, *as such*, could not create anything, just because creative effectiveness would introduce a more-than-ethical element; but in the same strong sense of "as such," no cow as such—no cow simply *qua* cow—could be brown, for brownness is more than cowness. Yet some cows are brown; ideas which we must distinguish may nevertheless correspond to aspects of some unified reality. When a suggested reality is *ethical and creative* marked-out-ness, we can reasonably theorize that its creative side is, like mass, heat, or color, able to exist only as an aspect of a greater whole (describable as *creative ethical* marked-out-ness). For first, perhaps only the addition of an ethical aspect will give plausibility to a relation of marking out for existence whose mere possibility has creative importance prior to the actual existence of anything. Second, if (as will be argued) there is some evidence that the world's general plan answers ethical requirements to a remarkable extent, then viewing this as mere chance would be almost as strange as believing that chance held together, at a thousand times and places, the weight and warmth and brownness whose coherence inspires talk of those many-sided entities, cows.

The two-sided relation between our good world's nature and existence is supposedly unified yet composite somewhat as a relation in which a line can stand to the rectangle which it cuts across: the relation of *bisecting (equally) in length in a way which*

also bisects in area. Note (a) the strong analogy between the two senses or varieties of bisection, preventing a pun; (b) that this analogy suggests, without itself guaranteeing, that the bisection in length is bisection in area also; (c) that though it is conceptually possible to cut the relation into two, there is no clear need to do so. However, this latter relation is provable. What status has extreme axiarchism's all-creating relation, and how will that status affect our readiness to believe in it?

The axiarchist can pursue the strategy outlined earlier, in connection with purely ethical relations, giving his creative relation the same status as the relation of similarity, whose reality is entirely bound up with the fact that the related terms are what they are. But whereas similarity fairly obviously has that status, there is nothing obvious with the axiarchist's relation. If such a relation connects the nature and existence of a best possible world, then with that unobvious fact the axiarchist story ends. We cannot gain further insight into "the mechanics of" the fact, for there are none, just as there are no "mechanics" behind the relation of similarity.

There may well be a limited analogy between being ethically designated for existence and being destined for it by physical forces. (If not, such phrases as "ethically marking out for existence with creative effectiveness" are jokes of the "arrived in tears and a sedan chair" class.) Moreover, the limited nature of the analogy—the fact that ethical needs are often ineffective—is no argument against axiarchism, since an ethical need might be ineffective because overruled by another: see discussion of the room left for moral struggle even in a best possible world. Still, metaphysics inspired by limited analogies has a limited ability to convince, and a reasonable man might perhaps find it 100 per cent unconvincing. There is no logical compulsion to look favorably on the final element in the extreme axiarchist's chain of explanation, for the ethically requisite is not logically requisite. . . . *And yet the alternative—the theory that the presence of the world, or of a creating deity, is a reasonless affair—equally asks us to believe something finally inexplicable*. Its correctness is logically possible, but how could we prove that a man should feel even a one per cent pull toward it?

Again, although ethical needs, however independent of the prior existence of any person or object, might lack all creative influence, the creative power attributed to them may be considered less "casual," less "brute," than any mere fact

that a world existed reasonlessly. Let me explain. As discussed, certain relations follow inevitably from the characters of the related terms. Similarity is one such; intrinsic value may be a second. Logic, I now suggest, provides others. Whether or not logical necessity is somehow relative to human wishes, in one aspect it is absolute. Once a logician's wishes have set up rules for passing from one set of words to another, additional wishes cannot be introduced to decide what shall follow from those rules. Compare chess. *The combination of chess conventions with certain chess-board maneuvers makes necessary Black's checkmatedness.* Similarly, *acceptance of various logical regulations together with certain premisses makes necessary a given conclusion.* The reality of these relations of necessitation is entirely secondary to that of the related terms, and not to additional human desires or other realities.

Such relations set up necessities of structure: the structure of logical proofs and board games, geometrical diagrams and jigsaw puzzles. For instance, given rules about what is meant by nuts, bolts, and twists, it follows necessarily that a nut can be twisted onto its bolt irrespective of which face is presented to the bolt. This structural necessity is in the world, not just a consequence of our rules for description, for it simplifies nut and bolt jobs. We could, however, play the logical game of reordering facts symbolized by words, the structural necessities revealed by the game mirroring this nut-and-bolt structural necessity (much as structural necessities governing the alignment of true furniture passing through true doors could be mirrored by those governing doll's-house furniture and doors).

Three points to be kept in sight. (1) If these and other relations have a reality entirely secondary to that of their terms, then they *must* connect those terms, making it odd to question further why they connect them. Reminder: such relations, involving no elements like chains existing apart from the trucks they link, cannot come and go as chains can. (2) It can appear equally odd to ask *why* the relations are secondary in their reality. (3) Some are relations of necessitation.—Now, extreme axiarchism can explain the world with a relation of this last class: one not of logical but of creative ethical necessitation. No one can be compelled to believe, for no logic can establish the secondary reality of the relation. But once again, if this alleged secondary reality is indeed a fact, then it is odd to ask *why* it is one.

Analogies between ethical and effective necessities give some reason for believing that there

might be a creative relation with an ethical aspect. The same end is furthered by arguments that ethical factors are themselves relational in type, and would have importance even as mere possibilities prior to the existence of all objects. What these arguments maintain is that an axiarchist cow (ethical necessity) is the sort of thing which might be brown (creatively powerful), in contrast to such obvious not-browns as the number 50. Other arguments suggest that the cow is indeed brown; arguments like, Only brown cows eat dandelions and the dandelions have vanished; arguments (see below) from scientific regularity and the presence of life. But once more, those fail to explain *why* the cow is brown. When the axiarchist has explained what he means by the creative effectiveness of value—reaching the alleged fact of a creative relation with an ethical aspect and wholly secondary in its reality to its terms—then to demand explanation of this fact of secondary reality can look like demanding why similarity in shape is a relation getting its reality solely from the related objects. The answers would be that these, being matters of secondary reality, could not be varied: that those are *the sorts of reality* which these relations are: which is more rebuke than explanation.

The eternal invariability of such relations gives some excuse for naming them "not casual and brute," but I do not insist. What is significant is that some concern simple facts such as *could not* be further explained. Some minds come to rest only when reaching facts like these. Maybe there is no reason for the world; yet some are discontented here, convinced that it might without absurdity be explained. Only an extreme axiarchism, with its ultimate reference to a fact of secondary reality which is by its very nature simple and inexplicable, will remove their discontent. (*Complex* secondary realities can in a way be explained: thus a complex similarity of shape, or logical necessity, can be considered the sum of many simpler ones, whose secondary reality is more obvious. That technique is used in logical proof, but is inapplicable to the alleged creative relation, at least in its creative aspect, which is what is in dispute. We might, of course, apply the technique to the ethical aspect, splitting the world's goodness into various ingredients.)

IX. THE ARGUMENT FROM SCIENTIFIC REGULARITY

Those unimpressed by the world's mere existence may yet ask why it bears an orderly pattern.

What guides events into a scientifically regular course? Compare geometrical regularities. We take any paper figure with four straight sides; mark and cut off the corners; discover that the marked angles exactly encircle a point; and straightway assume the presence of necessitating factors. Is it, then, likely that the most basic uniformities of physics lack a reason? Arguments from analogy are never conclusive, but here there appears room for a strong one. Admittedly we cannot explain physical regularities just as we do those of geometry, which are governed by logical necessities; yet why assume that the only necessities available are logical?

Some protest that the search for explanations presumes a self-explanatory tendency for events to obey laws of chance. This, they urge, involves confusions over the difficult term "probability." Governance by the rules of the mathematical theory of probability would be as much and as little awesome as governance by any others; for, as casinos know, it is difficult to construct devices which illustrate those rules precisely. Shall we imagine an imp twirling his roulette wheel to give all logical possibilities an equal likelihood, and frustrated only by superior demons? However, the force of such protest can be much reduced. What makes physical order remarkable is that, of all conceivable ways of rearranging events, very few, or none, would yield rules as simple as those of their actual arrangement. Now, to fake logical proof of the impressiveness of such facts, one needs confusions over terms like "probability"; yet there is no evident need for logical proof, and hence no need even to mention such terms. If the mere contrast, between the few possibilities of order and the endless possibilities of disorder, fails to impress, then no maneuvers with the word "probable" will improve matters; and if a logical certificate is demanded for every thought jump, we shall do precious little thinking.

Experience might be thought to indicate a tendency for mathematical laws of probability to be obeyed, except when there exist countermanding necessities. No doubt it is hard to balance the perfect roulette wheel, but at least we understand why a good wheel is good. Casinos try to ensure that the known laws of nature have no interest (like the interest of gravitation in the usual failure of coins to come to rest on their edges) in the appearance of any particular one among the outcomes which interest gamblers. Given a lack of interested laws, "chance" results are predictable. We need not test each regular solid in turn, to be sure that all can be used as dice: physics could have forecast this if

dice had never been thrown. Neither do we require specific tests to show that laws of chance apply in experiments on card-guessing, to be startled by any failure of those laws. Logic allows the world to be a rag-bag of disconnected regularities, but it is not.

What excites interest in scientific uniformities is not precisely the failure of all logical possibilities to be realized equally frequently—for why not expect a tomato soup world, everywhere identical? Rather, the interest is in the relatively simple pattern into which various possibilities fall. If unsure how accurately a coin is flattened, we have small cause for surprise if it lands more often Heads than Tails, but what if the Heads and Tails yield a Morse-code guide to the laws of nature? The question of why the world obeys such laws invites appeal to the ethical need for there to exist an orderly world. Contrast unintelligible appeals to "secret powers" inside objects, giving them "causative influence" over one another.

X. ARGUMENTS FROM THE PRESENCE OF LIFE

Scientific order has value as being essential to life. A promising axiarchist strategy is to argue that many details of the order are likewise essential. Such an "argument from design" need not point toward a designing God if ethical requirements might themselves be creatively effective, since it is odd to guarantee a role for a deity by adding: just so long as they require nothing too complicated.

The argument can survive evolutionary theory, as the more we demonstrate life's inevitability in a universe obeying the laws which ours does, the more awesome we may find those laws. Books on the subject are given to listing the roles of various chemicals; almost every element is essential to life; but there are many simpler truths equally worth attention. Thus, imagine the effect of laws yielding radical changes in properties with each movement. Luckily the laws we see favor stability—particularly the stability of life's constant eddy in a stream of ever-different atoms. Again, consider the principles controlling distances between bodies; gravitational forces forming nebulae, stars, and earth; those balanced by centrifugal ones, maintaining that most beautiful planetary system which Newton considered planned by an intelligent and powerful being. A further balance inside atoms makes them building bricks far more versatile than the musket balls of an early physics, joining them with the reversible physical and chemical ties essential to living systems, as well as allowing for

the steady release of nuclear energy (in suns) to provide the ultimate source of power for those systems.

In contrast are more bizarre truths. In quantum physics, what appear to be waves—and hence pass energy into patterns whose intricacy sets the stage for life's complexities—give up this energy in un-wavelike bursts at points predictable only statistically. Such concentration of apparently dissipated energy permits photosynthesis, vision, and so on. Again, in a universe seemingly expanding fast, it is providential that no matter how rapidly a living system moves relative to others, its internal development can proceed identically. This is remarkable, as forces mediating interaction between any such system's particles are propagated through the space separating those particles at a finite speed: one not high by comparison with the estimated speed of recession of the farthest visible objects. The old philosophical realization that motion is relative does not explain such "relativistic" effects: it does not predict that electromagnetic waves should have the same measured speed relative to us, however we move. The interplay of physical principles involved is far from simple, and the result, that the parts of living bodies would continue to cooperate when moving very fast relative to any center of expansion which the universe may have, is fortunate.

One could continue in this vein, but the point is made. Life could be thought to balance on such a razor-edge that it must have been put in balance there.

A counter-argument is that life could never provide evidence of a rule of value, for, were physical laws not such as to favor life, we should not be here to discuss them. Now, I admit that, since we are here, it is not remarkable, in view of *that*, that the world can support life: yet it may be remarkable that we are here. That life should *never* strike us as remarkable enough to require metaphysical explanation involves the absurdity that, even could we know that life would have appeared only if a demon had thrown a million Heads in a pre-specified million tosses, *still* life's presence would not prove the coin double-headed.

XI. SPECIMEN BELIEVERS

(A) Though many accept some kind of rule of value, fewer develop the extreme axiarchism which explains all objects. Leibniz, I think, does. Unfortunately, as a pioneer logician he is a target for charitable reinterpretations of metaphysics as for-

givable logical error. Thus his publicly expressed optimism—the *Theodicy* includes sophisticated defense of moral struggle in a best possible world—is dismissed in favor of an allegedly quite separate "secret doctrine," which holds that the world is selected by a dry logical necessity for as many existents as possible. Such reinterpretation, besides giving him a silly theory, overlooks his insistence that even small volumes contain infinity existents. Again, it forgets the philosophically traditional association of "amount or richness of existence" (not a simply numerical idea) with degree of value. It is surely more charitable to make his axiarchism fundamental, classing his trouble as not distinguishing a logical affair, namely, what *could usefully be written into the concepts* of things, from a metaphysical one, namely, what perhaps *follows from the natures* of such things (for that may contain extras, such as creative ethical requirements). We still struggle with this distinction.

Much disputed is the part he gives to God. The logic-gone-wrong interpretation makes him propose a dryly logical war of essences struggling to exist, the outcome not depending on God who is secretly considered non-existent. A less fantastic picture, which saves his writings from dishonesty, is that God, though just one element in a cosmos existing for reasons of value, is the most prominent element: thus the others have characters largely determined by how they can best interact with him. This picture, which explains God's influence much like that of other things, is suggested by such statements as that "unless in the very nature of essence there were some inclination to exist, nothing would exist," or that the world will be as rich as possible "if it is once given that being is superior to non-being, even if there is no further determining principle." That Leibniz did not develop it clearly might reflect the feeling of his age, still surviving today, that it was somehow blasphemous to make God's existence and power anything other than the source of all explanations.

(B) Descartes thinks it no courtesy to describe God's existence as reasonless. God is his own creative ground, for "I have not said that it is impossible for anything to be its own efficient cause, though that is manifestly true when the meaning of efficient cause is restricted to causes which are prior in time to their effects and different from them." Self-creation is associated with "perfection," a term with undeniable ethical overtones. But again there are problems in distinguishing bad logic from metaphysics. The Cartesian argument for God can

seem to attempt the impossible: use of words in a game whose own structural necessities mirror, not a necessity governing the structure of what *might exist* (such as guarantees the possibility of joining any three points with the arc of a circle), but an effective necessity of existence. However, his *Replies to Objections* admit that nothing follows from definition of God as perfect: "that a word implies something is no reason for that thing." Emphasis moves from words to a "clear and distinct" insight into how God's presence follows necessarily from his perfection. (It admittedly complicates matters that "perfection" means to him "completeness of being" as well as "supreme good"; yet God's necessary being is associated with the fact that a thing existing by its own power has "every property which we can imagine it would bestow upon itself," i.e., every property ethically desirable.)

Probably Descartes misinterprets the tie between perfection and necessary existence, not recognizing that being perfect would mean having a constitution *which necessitates existence ethically (and maybe effectively)*, and not *which includes effectively necessary existence as an element*. Still, his argument comes alive when considered not a proof but a pointer to a possible ethical reason for God's presence.

(C) Several "new theologians" reject a person named "God," and for every writer who attacks them for denying religion, another does so for unoriginality, pointing to the long theological tradition behind them. Much that is obscure is said by writers in this tradition, but some are most simply interpreted as viewing God-as-a-person as mythical personification of a force of ethical necessity. The ethical pull which originates and draws onward the stream of events is not imagined as that of any individual, either inside or beyond the world around us. Tillich's God is not a being but "the creative and abysmal ground of being," as well as "that which concerns man ultimately" (a reference to value). Robinson (formerly Bishop of Woolwich) makes no call for "the feat of persuading oneself of the existence of a super-Being"; "the word 'God' denotes the creative ground and meaning of all our existence" (where "meaning" equals "worth"). Remember, though, that when axiarchism is ambitious to explain all existence, one need not exclude God-as-a-person in this way, since such a person is himself in reach of its explanations.

XII. IMPACT ON OTHER BELIEFS

It is frightening how many strange pictures pre-

sent themselves, given the single assumption that things exist because they should.

(I) If we explain objects through their value, those with purely instrumental worth must go. For their value is only as a means to their effects: now, if it is logically possible that such effects were produced directly by ethical requirements, then that is what the thoroughgoing axiarchist will maintain. Axiarchism's razor begins with dismissal of the world in ages simply preparing the arrival of conscious states, for presumably only consciousness has intrinsic worth. (But descriptions of those ages retain interest, as describing the greater pattern which would round off the patterns of consciousness.) Where the razor stops depends on what is essential to conscious life itself. This reasoning seems to offer a reprieve for phenomenism (immaterialism).

(II) If conscious life is a mere pattern of electrical activity sweeping cerebral lamps whose sole connection relevant to value is that they display this pattern, then it would seem too much an abstraction, like the distribution of buzzes in a beehive, to be a strong candidate for intrinsic worth. (If beings scattered round the universe each controlled the discharges of a human brain cell transported to his planet, and if all then conspired to place those cells in a one-to-one relationship to those of some living man, so far as concerned their activity, then on the view considered we should have two activity patterns with equal claims to ethical desirability. That can look unpalatable.) I am not doubting physiological explanations of thought-patterns, but only certain strictly philosophical assumptions which scientists, trespassing, tend to make about the ontological status of whatever carries these patterns. In dealing with consciousness, axiarchism appears to need a theory allowing a unity of existence—defined as something logically capable of existing if all else vanished, and containing no parts similarly capable, since any parts which it has are only aspects or abstractions—to have a structure. Intricate patterning may be compatible with oneness of being.

(III) An extreme axiarchism requires an account of time's passage answering the question: If the world is best possible, why does it ever change? The picture of time as a space-like fourth dimension offers one such account. It presents variation with time as supplementing value—the worth of today being added to that of yesterday—instead of merely varying it. Time limitations on experience (limited life-span) become on a level with spatia

limitations (boundaries to one's visual field, say), and it becomes very odd to pity people of past centuries just because they are past.

Any patterns of experience allegedly disproving a four-dimensional block might equally be viewed as patterns developing along the time axis of that block itself, as a pattern of threads develops across a carpet. But perhaps axiarchism should not propose a *space-like* fourth dimension (as "block" suggests): "space-like" may be needlessly specific. It should perhaps limit itself to rejecting accounts which utterly rule out the possibility of time's being space-like: for instance, the account which has the present preying on the past for the stuff of its existence, so that today side by side with yesterday would be like two buildings made with one building's-worth of brick. If the value of states at all times is to be added, not averaged, in finding the value of the world's existence, those states must at least be available for placement along a space-like dimension: available in such a way that only addition of space-like relations would be necessary for such placement. A deity would not have first to re-create past states and create future ones, if wishing to place them on a line beside what is to us present. Whether it is simplest to class time-relations as themselves space-like is another affair.

(IV) The approaches of (II) and (III), dealing

with conscious unity and time, involve great conceptual difficulties. Let me complicate matters further. If combined, they give a picture particularly attractive to axiarchists: one allowing the value of a whole lifetime's consciousness to be added in finding the value of a unity of existence. (This suggests nothing so silly as that men of today can inspect and prophesy their doings tomorrow, for whether the pattern of a man's conscious life is present in its entirety in a unity of existence is an ontological question having nothing to do with the details of that pattern—presence or absence of prophetic ability being one such detail.) Moreover, once we allow complexity despite unity of being, there seems no means of setting limits to this complexity, so that the same unifying approach might be applied to the entire world of individual lives. (Once more, nothing so silly as that everyone can report on everyone else's experiences.) Finally, if ethical needs can produce one world, they would presumably produce any number of similar worlds. Now, if endless worlds of consciousness, each unified in its existence, look not impossible but incredible—too good to be true—then do not join me on the axiarchist slide. You have but to deny, as reasonable men certainly may, that ethical needs by themselves bring about anything.

University of Guelph

Received July 1, 1969

III. SOME APPLICATIONS OF INDUCTIVE LOGIC TO THE THEORY OF LANGUAGE

L. JONATHAN COHEN

PSYCHOLINGUISTS often take it for granted that induction typically estimates the probability of a thing's being, say, a β if it belongs to a statistical population of α 's, on the basis of the actual frequency of β 's in an observed sample of α 's. I want to point out, first, that this assumption about induction is an indispensable premiss of Chomsky and Miller's very powerful 1963 argument¹ for the existence of innate syntactic universals, and that a different conception of induction invalidates that argument and several other arguments for such universals; secondly, that a solution then emerges for the problem about how it is possible to understand semi-grammatical sentences; and thirdly, that an explanation also then becomes available for the possibility of classifying predicative expressions into different semantical categories that are in some way to determine whether, or to what extent, a syntactically well-formed string constitutes a semantically normal, or a semantically abnormal sentence. The overall purpose of tackling three such issues in a single paper is to show up the importance of inductive logic for an adequate theory of language. But what I have to say is quite independent of current controversies about how to write the grammar of a natural language and is not intended in any way to belittle the classical significance of Chomsky's work in this field.

I

Chomsky and Miller began their argument,² by assuming that an empirical, inductive language-learning device would treat grammatical utterances as Markovian processes, where a Markovian pro-

cess is understood, roughly, as being a sequence of events such that at any durationless point p all the information about the preceding history of the sequence that is relevant to determining a probability for the next event after p is given by the outcome immediately preceding p . In effect, therefore, they assumed that learning the grammar of a language by induction must consist of making a very large number of estimates about the probability of an expression's being such-and-such if it belongs to the population of expressions that are immediately preceded by so-and-so. They then pointed out that in any case a language-learning device of this kind could at best extract only the set of grammatical sentences that are of not more than some definite number of words or morphemes, and they rightly objected, on the assumption that grammatical sentences can be of any finite length whatever, that a Markovian language-learning device would produce a knowledge of a language that failed to embrace infinitely many of the grammatical sentences that are actually available in the language. But even if a sentence never had more than 15 words and the range of choice, in any position in a sentence, were restricted by the categories of the preceding words to a word of one or other of, say, four categories, the number of probabilities to be estimated would be 4^{15} . So a child would have to learn the values of approximately 10^9 parameters in a childhood lasting approximately 10^8 seconds (i.e., three years).

What this argument actually establishes is that no inductive language-learning device which works by the estimation of Markovian probabilities can constitute an adequate model for the language-learning capacity of a human child. It is thus a powerful move against those numerous writers who

¹ N. Chomsky and G. Miller, "Finitary Models of Language Users," *Handbook of Mathematical Psychology*, ed. by R. D. Luce, R. Bush, and E. Galanter, vol. 2 (1963) p. 424 ff., esp. p. 430.

² Chomsky has also published several other, less rigorous arguments for innate syntactic universals. One of the latest is to be found in Noam Chomsky and Morris Halle, *The Sound Pattern of English* (New York, 1968), p. 4. I am concerned here, however, only with the case for innate syntactic universals, not with that for innate phonological universals, which may well be stronger.

reject innate syntactic universals on the ground that some general learning strategy could accomplish the task of language-acquisition, but do not go on to specify the structure of such a strategy. Nevertheless, Chomsky and Miller's argument does not establish that no general learning strategy of any kind could do the job, and what follows is an attempt to specify the structure of an inductive learning strategy that is not open—in respect to childhood language-learning—to the Chomsky-Miller argument about too little time.

Other things being equal, one estimate of the probability that an α thing is a β is judged preferable to another if it is based on a larger sample of α 's, for then there is a greater probability (of a higher-order kind) that the frequency of β 's in the sample will match the frequency or probability of β 's in the population as a whole within any chosen limits of approximation. So this mode of reasoning may be regarded as a generalization of what Bacon and Mill called enumerative induction, since, other things being equal, it is the sheer number of evidential instances—i.e., the size of the evidential sample—that makes one enumerative induction better than another. The difference is just that whereas enumerative induction was characteristically conceived to use this basis for trying to establish universal propositions, i.e., propositions to the effect that the probability of an α 's being a β is 1, the concept of an inductive estimate is not restricted in regard to the size of the probability with which it is concerned and it thus constitutes a generalization of the notion of enumerative induction. But, though Bacon and Mill are archetypal empiricists in epistemology, neither of them regarded enumerative induction as the main way to gain general knowledge from sensory experience. They emphasized instead the value of variety, as distinct from mere multiplicity, in experimental instances, as in what Mill called the methods of agreement and difference.

Accordingly Chomsky and Miller's assumption, that an empiricist theory of syntax-learning must be confined essentially to the procedures of enumerative induction, cannot be related to the empiricism of Bacon and Mill. Instead it stems from an identification of empiricism here with the associationist tradition that goes back not to Bacon but to Hume. For in his characteristically sceptical passages (as distinct from his brief section in the *Treatise* on the "rules by which to judge of causes and effects") Hume paid no attention at all to the salient features of experimental reasoning as Bacon

had conceived it. It is not just that Hume preferred to offer here a causal explanation of our general beliefs about matters of fact rather than a logic for them. He was also curiously narrow even in the type of cause that he was prepared to recognize. He talked, like a proto-Markovian, about the effect of constantly observing one kind of event to succeed another, not about the effect of observing one kind of event to succeed another in a variety of different circumstances. He seems pre-occupied, and Chomsky and Miller likewise, with the kind of causal influencing of the mind that is analogous to what Bacon called induction by simple enumeration, and ignores altogether the kind of causal influencing that is analogous to induction by variation of circumstance. But there is no plausible reason why the title "empirical" should be withheld from the latter method of learning, if it is granted to the former.

Indeed, if science in fact prefers variational induction to Markovian learning, there is a strong presumption that the former is a better method than the latter for accomplishing complex learning-tasks. Now this is very important in relation to the problem of language-learning. For if someone wishes to show that it would take children too long to learn linguistic syntax empirically he must show this to be so even if they use the most effective learning strategy. But I shall try to show that variational induction—on an appropriate construction—can in principle produce results in a way quite different from Markovian learning, and that therefore the Chomsky-Miller argument against empirical language-learning is invalid. (I shall not attempt to argue the thesis that variational induction would actually produce results fast enough for childhood language-learning. To prove that thesis, or to disprove it, we need to know a very great deal more than we do at present about the structure of the syntax that has to be learned and about the effect of different cultural and socio-economic conditions on the various stages of language-learning.)

We must distinguish here between inductive logic and inductive heuristics, as Bacon and Mill sometimes failed to do. Roughly, the standard subject-matter of inductive logic is a dyadic support-relation between propositions, that is timeless though a matter of degree, while the standard aim of inductive heuristics is to enunciate a set of rules whereby a man may discover which propositions constitute the answers to his intellectual problems that enjoy the best inductive support

from propositions of immediate evidence.³ So, though what is directly in point here is the heuristic methodology of induction by variation of circumstance, this methodology is bound to rely on an implicit theory of inductive logic. I shall sketch such a heuristic methodology briefly here, but for the underlying logic, and its elaboration and defense in detail, I must refer the reader elsewhere.⁴

We have to suppose, in any particular field of research, first of all a certain set of materially similar universal hypotheses, where material similarity is defined over a subject-matter—i.e., defined by reference to the non-logical vocabulary of the hypotheses concerned—and secondly, a set of natural variables (or variables, for short) that are inductively relevant to such hypotheses. By a natural variable here I mean a set of observable circumstances—called the “variants” of the variable—the descriptions of which are incompatible with one another.

A two-variant variable may be said to be inductively relevant to a set of materially similar hypotheses if and only if at least one of these hypotheses has been falsified when one variant of the variable was present, and has held good when the other variant was present, other circumstances being the same in both cases. An n -variant variable, where $n > 2$, may be said to be inductively relevant if and only if each of its variants is a member of some two-variant variable that is relevant.

To test a hypothesis—a logically contingent, universally quantified, conditional—we seek to discover whether it holds good under all possible combinations of variants of certain of the relevant variables, and the more relevant variables that are manipulated in the test the greater the support—*ceteris paribus*—that is forthcoming from a favorable test result. To ensure maximum comparability,⁵ however, we need to establish a single hierarchy of cumulatively more and more thorough tests, on the basis of a well-ordering for the set of relevant variables; and it is appropriate to order the list of variables that are relevant for hypotheses of a particular type according to the known frequency with which variants of them have been successful in

falsifying hypotheses of that type (and alphabetically, say, where this procedure produces a tie.) We must also preface this list by an appropriate subset of any variable (or variables) that is mentioned, or has a variant that is mentioned, in the antecedent of the hypothesis to be tested. For, if the hypothesis is a second-order generalization asserting a correlation between two or more natural variables (as between the period for which a body has been falling on to the earth, say, and the velocity at which it is falling,) the most relevant factor to vary in a test is indicated by the antecedent of the hypothesis, though where that variable has a continuum of variants the test has to be conducted over an appropriately chosen subset of variants. In terms of this ordering of relevant variables the least thorough test, t_1 , seeks to establish whether the hypothesis holds good in each variant of v_1 , where no variant of any other relevant variable affects the situation. The next more thorough test, t_2 , seeks to establish the same for every possible combination of v_1 and v_2 ; . . . and t_n for every possible combination of $v_1, v_2, v_3, \dots v_n$.

The grade of inductive support that has been obtained for one hypothesis may thus be ranked, in comparison with that obtained for another within the same field of research, in terms of the thoroughness of the respective tests that each has passed; and also the grade of inductive support that is provided for a given hypothesis by one set of test-results may be compared in similar terms with that which is provided by another. If a hypothesis has passed the most thorough test that we know how to construct, then, so far as we know, it has full inductive support. If it has failed test t_1 , it has zero-grade support. If it first fails test t_{i+1} it has just i th grade support.⁶ Our series of tests $t_1, t_2, \dots t_n$ enables us to map the support given to a hypothesis H , by a proposition reporting the results of tests of H , on to the first $n + 1$ integers, $n \geq 0$. Indeed, we can grade in this way not only the support that exists for ordinary first- and second-order generalizations, but also the support that exists for third-order systems of postulates, or “theories,”—involving concepts of a higher level of abstractness

³ Because Bacon recommended scientists to collect evidence before they generalize, philosophers have sometimes used the word “inductive” in a sense that makes it apply only to methodologies incorporating this recommendation, and for such philosophers the term “inductivist” has thus acquired a pejorative connotation. But when looked at in terms of the inductive/deductive contrast the fundamental type of relation at issue is a logical, not a temporal one; and if “inductive” is used in a *logical* sense Bacon must be praised, not blamed, for his inductivist analyses.

⁴ L. Jonathan Cohen, *The Implications of Induction* (London, 1970).

⁵ The argument is given in detail *ibid.*, §§ 5–7.

⁶ If failure of test t_i is combined with success from test t_j , where $j > i$, a hidden variable must be operating: cf. *ibid.* § 8. Hempel’s and Goodman’s paradoxes do not arise: cf. *ibid.*, § 11.

—from which very many such generalizations are deducible.⁷

It is important to note that this mode of inductive reasoning offers us a systematic method of modifying our hypotheses in order to circumvent their falsification by unfavorable test-results. For example, suppose variation from arctic to temperate zones was considered the least relevant variable for hypotheses about fur-color, and thus figured only in the most thorough test of such hypotheses. Then, if the hypothesis

Anything, if it is a hare, is gray

passed all the other tests, and failed the most thorough one, just because arctic hares are white in winter, the hypothesis could easily be modified so as to pass even the most thorough test. We just need to modify it, so that it reads

Anything, if it is a temperate-zone hare, is gray.

Insertion of the qualification "temperate-zone" here excludes the possibility of testing this version of the hypothesis in relation to arctic hares and *a fortiori* excludes the possibility of its failing such a test. In general any report of test-results that gives less than full support to a hypothesis may be said to give a suitably modified version of that hypothesis full support, where suitable modification is understood to consist in the insertion of a proviso of inoperativeness into the antecedent of the hypothesis for each relevant variable that was not manipulated in the test reported and the insertion of a mention of a non-falsificatory variant for each variable that was manipulated in the reported test but has a variant that contributed to falsifying the hypothesis—or in the insertion, for each such variable v_1 , of the proviso that no variant of v_1 is exercising any effect.

It follows that the more we modify and complicate our hypothesis, in one of the ways described, the smaller is the number of evidential instances, or individual experiments, that we need to obtain full support for a modified version of the hypothesis. E.g., if each of our five relevant variables has just three variants, the number of evidential instances required for full support reduces from 81 in the most thorough test, to 27 in the next most thorough, then to 9, then to 3, then to 1.

We can thus conceive of an inductive language-learning device not as a compiler of Markovian probabilities in relation to a predetermined list of

categories, but as a formulator and tester of hypotheses. The occurrence of a successful test-result may be evidenced either by observation of other speakers' favorable reactions to the device's own attempts at speech in accordance with its hypotheses, or by the observation of other speakers' utterances that conform to these hypotheses. Typically such an inductive device would profit by over-generalizing in the initial construction of its hypotheses, as, e.g., in the hypothesis that any string of the form noun-verb-noun is grammatical in English. Then, when these hypotheses come to be falsified—in the sense that adverse adult reactions appear to some utterances of the form in question, or no adult utterances are observed to instantiate some of its subforms⁸—a store of relevant variables can gradually be amassed: e.g., the singular-plural variable for nouns preceding verbs, the singular-plural variable for verbs, the transitive-intransitive variable for verbs, and so on. In terms of variants of these variables it becomes possible to describe heavily qualified hypotheses that can be seen to derive full support from relatively few instances. Moreover, even if the process of learning involves making many initial mistakes, none of these need be altogether false starts, but merely mistakes of over-generalization. If an inductive learning-device discovers it has made such a mistake and recognizes the relevant variable that is responsible, it will not just reject the over-general version of its hypothesis. It will also adopt an appropriately qualified version for future tests, thus taking a positive step toward obtaining a fully supported version.

Admittedly, if each hypothesis to be tested were selected at random from an indefinitely wide range of possible hypotheses, an inductive learning-device could not be expected to make relatively steady progress toward discovering a language's syntax. But if the device always hypothesizes as boldly as it can about its initially noticed data, and starts with relatively few concepts at its disposal, and acquires more only by observing the circumstances in which its over-generalizations are falsified, and wherever possible, prefers to qualify and re-test a falsified hypothesis rather than reject it outright, and can introduce only a certain degree of clause-nesting into its hypotheses until all hypotheses of this degree have already been eliminated, the device will make much better progress.

Again, if the device has always to look at what

⁷ Cf. *ibid.*, § 9.

⁸ Cf. E. Mark Gold, "Language Identification in the Limit," *Information and Control*, vol. 10 (1967) p. 453 f.

happens in each combination of variants of relevant variables to know what modifications, if any, can be introduced to guarantee a hypothesis against falsification, then, if there were even as many as 10 relevant variables and each had 10 variants, apparently 10^{10} evidential instances would be needed, and the prospects for variational induction would scarcely look brighter than those for Markovian learning.

But there is a way of obtaining the same result from far fewer instances. Essentially this method consists in conducting a series of separate tests in relation to each relevant variable, instead of one big test in relation to all of them at once. Where n mutually independent variables are relevant, and each has m variants, this shorter method requires under the most favorable conditions nm , not n^m instances. But how can one make it safe to interpret each trial-outcome—each evidential instance—as showing how just one variable v_1 , and no other, bears on the hypothesis? How can one guard against the possibility that an unsuccessful trial-outcome was due to the presence of some variant of another relevant variable, v_j , so that the hypothesis requires to be modified in terms of v_j , not v_1 ? In natural science it is sometimes possible to screen off unwanted types of causal influence from an experimental setup. A substance may be purified, a lead shield erected, a vacuum introduced, and so on. But in syntax-learning this is not so often possible. Sometimes, e.g., a singular/plural variable might be screened off, as it were, by the use of a word that had the same form for both singular and plural. More normally we must suppose instead that a syntax-learner achieves an unambiguous interpretation of each unsuccessful trial-outcome in a single-variable test by attending either to an immediate utterance, by his adult hearer, of a corrected version of his own grammatically deviant string, or to some such utterance by an adult speaker on another, otherwise similar, occasion.

I.e., perception of a difference between the syntax-learner's utterance and some adult speaker's utterance may be supposed to reveal at least one of the respects in which the former was deviant. So from the same piece of evidence the concept of a particular variant of a particular relevant variable may be acquired, if it has not been acquired already, and also one of the syntax-learner's hypotheses may be appropriately modified. In practice the evidential instances for a test in relation to some one particular relevant variable may occur on different occasions over a considerable period of time. But, given a suitable filing system for his evidential data and suitable cooperation by adults, a syntax-learner would require at most $n(2m - 1)$ evidential instances to acquire an appropriately modified version of a syntactic hypothesis—i.e., a version that *would* enjoy full support if it were subjected to the most thorough possible test.

Nor should an inductive language-learning device of this kind be thought to be confined to the formulation of the most elementary type of hypothesis—to listing labeled bracketings of non-deviant surface structures. As soon as it acquires a stock of relevant variables, we must expect its nisus toward generalization will lead it to formulate hypotheses about relationships between these variables which will subsume and explain, as it were, the more elementary hypotheses that have already been established: e.g., "The singular/plural variable for verbs varies directly with the singular/plural variable for preceding nouns," or "The insertability of a *by* . . . phrase after the verb varies inversely with the active/passive variable."⁹ Finally, we may expect yet higher-order generalizations in which theories are tested that subsume and explain these correlations by asserting the truth of whatever follows from a certain set of postulates (e.g., a theory asserting the grammaticality of any string generated from a given lexicon by a certain set of phrase-structure and transformational rules). An

⁹ Perhaps someone will ask how there could still be so many unsolved problems in natural science if each of us were in fact born with an all-purpose inductive mechanism that accomplished such an impressive task as that of tacit syntax-learning in so short a period as childhood. Part of the answer to this question is that in syntax the number of relevant variables is vastly fewer, and their size vastly smaller, and hypotheses are much more easily tested, than in most fields of natural science. The other part of the answer is that when an engineer's pioneering design for a supersonic airplane, say, is faulty because his aerodynamic theory is incorrect, nature does not supply him with an opportunity to inspect a viable supersonic airplane and discover its distinctive features, in the way that adult speakers constantly provide a child with paradigms of correct speech from which he can learn to correct his faults.

Rather more difficult to answer is the question why an adult native speaker cannot write a grammar for his language as easily as he internalizes such a grammar when a child. But this question is presumably an empirical one, which it is for psychologists or neurologists to answer. Nothing at present known seems to suggest that an answer will be more readily forthcoming if we suppose the innateness of a specific strategy for syntax-learning than if we suppose the adequacy of a general learning strategy for this purpose.

inductive language-learning device no more requires a distinct portion of evidence for each sentence-frame—i.e., each pattern of grammatical sentence, as Chomsky and Miller suppose—than Newton required a distinct portion of evidence for the operation of gravity between material objects at each possible distance of these objects from one another and at each possible ratio of their masses to one another.¹⁰

Four further points, at least, need to be made.

The first is that there is obviously a problem about how, and in what form, an inductive language-learning device might acquire the concepts in terms of which it could construct its syntactic hypotheses. We may suppose the logical machinery innate—i.e., part of the device itself. But this machinery is applicable indifferently to causal learning, ethics-learning, and so on. How can one explain the employment of specifically syntactic concepts, such as those of “noun” or “verb” or “cyclical application of a transformational rule”? It is not difficult to suggest a possible way in which a child might acquire the concepts that are necessary for generalizations about surface structure. The multiplicity of concepts involved can be reduced to two primitive ones (“sentence” and “nominal”) by assuming a mode of derivation for the others like that used in categorial grammars.¹¹ Then the acquisition of these two primitive concepts has to be supposed to stem, as to “sentence,” from experience of utterances formed by varying recombinations of a stock of recognizably recurring components, and, as to “nominal,” from experience of the utterance of some such components in isolation from others (in the presence, no

doubt, of what Quine called “conspicuously segregated objects”).¹²

So far as the more abstract concepts of transformational grammar are concerned, it seems rather unlikely that this mode of derivation is possible. But it does not follow that we must therefore postulate the innateness of certain specifically linguistic concepts, if we are to suppose that language-learners at some stage internalize a rather small set of very general and highly abstract rules, which stand to their initial hypotheses and correlations much as a scientific theory stands to the causal uniformities it explains. For it may well be that the highly abstract concepts requisite at this level are all capable of non-linguistic as well as of linguistic realization. It is certainly not obvious, for instance, that the ability to operate grammatically in terms of recycled applications of transformational rules is substantially different from a child's ability to transform a horizontal line of playing cards by adding pyramidally supported tiers of playing cards, or that the structure-dependence of, say, an interrogative transformation is substantially different from the structure dependence of an individual jump in the children's game of hopscotch. Indeed as the structure of theories of transformational grammar becomes more and more sophisticated, and the concepts they invoke become more and more abstract—partly, at least, to cope with the growing volume of transformational accounts of different languages—less and less hope is justified that we shall establish in this way the innateness of certain specifically linguistic concepts. The more abstract the concepts that are invoked, the more plausible it is to suppose that, if

¹⁰ Empiricist theories of syntax-learning have sometimes been criticized on the ground that they assume the syntactic structure of a sentence to be expressible by a sequence of markers representing the classes of which the words comprising the sentences are members, e.g., T. G. Bever, J. A. Fodor, and W. Weksel, “A Critique of Contextual Generalization,” *Psychological Review* vol. 72 (1965) p. 467 ff. The proper answer to such a criticism is to emphasize the continuity of inductive methodology from first-order, elementary generalizations, through second-order, correlational generalizations, to third-order, theoretical generalizations. In other words, when the term “empiricist” is used to describe a theory about how knowledge is acquired, as distinct from how it is validated, one must recognize not only that every empiricist has to postulate some mechanism for the acquisition of knowledge, but also that association is far from being the only type of mechanism that an empiricist can postulate. So the choice is far from being just between Humeianism, on the one hand, and Rationalism on the other. The real issue is about the scope and nature of general learning strategies as against specific ones. Indeed the much-bruited dichotomy between Empiricism and Rationalism is not only an oversimplification on the empiricist side. It is also historically inaccurate in relation to the classical Rationalists. So far were Descartes and Leibniz from supposing that innate ideas explained our actual linguistic competence that they expressly asserted the defectiveness of existing languages in relation to the true, innately grounded structure of science. What the latter required on their view was a new language that would become genuinely universal, and they both proposed the construction of such an artificial language: cf. L. Jonathan Cohen, “On the Project of a Universal Character,” *Mind*, vol. 63 (1954) p. 49 ff.

¹¹ Cf. Yehoshua Bar Hillel, *Language and Information* (1964), p. 99 ff.

¹² The early presence of some such concept as that of a nominal (however it be obtained) is evident from well-known investigations into the speech of two- to three-year old children. Cf. D. McNeill, “Developmental Psycholinguistics” in Frank Smith and G. A. Miller (eds.), *The Genesis of Language* (Cambridge, 1966) p. 15 ff., where the term “open class” is used for this concept.

innate at all, these concepts represent certain general abilities that have indefinitely many applications. After all, most evolutionary explanations of innate abilities are in any case rather speculative. But the task of explaining the innateness of certain specifically syntactic principles, in terms of Darwinian evolution, is in principle a great deal more difficult than that of explaining the innateness of certain more general abilities, and one would be justifiably reluctant to sacrifice Darwinian conceptions of evolution unless or until a better theory replaces them. Moreover, it is worth remarking that, if the deep theoretical systematization of syntactic uniformities did in fact involve the use of concepts, models or structures that have a more obvious, down-to-earth application outside the field of grammatical investigation or grammar-learning, there would be a striking analogy here with the familiar way in which physical theories of light, say, or electricity commonly employ some everyday concepts, like "wave" or "current," in a special, technical sense. When regarded in the light of this analogy it might seem a little surprising if the theoretical, or quasi-theoretical, models needed for the systematization of linguistic syntax did not enjoy any structural isomorphism with everyday non-linguistic concepts.

A second point which needs to be made here is that inductive reasoning of the kind described can often issue in conclusions that are later found to be erroneous, even when all observations and descriptions have been correctly made. Judgments about how much one proposition supports another are, on the present view, contingent and corrigible, not analytic and *a priori* as in most contemporary confirmation theories (e.g., Carnap's). Characteristically what leads to an error of inductive assessment is that a hidden variable has been contributing to some of the test-results. In such circumstances it is possible to suppose a hypothesis fully supported which subsequently turns out false, or to be faced with different results from two performances of the same test. One may then be ignorant of how to construct a suitably modified version of the hypothesis that could obtain full support. This is a familiar type of situation in the natural sciences. For example, failure to detect the relevance of the pregnant/non-pregnant variable to certain hypotheses about non-toxicity turned out to have been partly responsible for erroneous conclusions about the safety of thalidomide. So it is easy to account for

the fact that any inductive language-learning device is bound to make mistakes on its way to acquiring mastery of a language. For example, it may, for a while, miss the relevance of such morphological variables as the strong-verb/weak-verb variable in English, and produce ungrammatical sentences like **The man singed a song*. Sometimes, too, the inductive relevance of certain psychological variables may be missed, so that, for example, an ellipse, say, or an aposiopesis is wrongly taken as an evidential instance for some hypothesis about what is grammatical, whereas it is in fact evidence only for what is intelligible or acceptable. But it is a special virtue of induction from variety of circumstance as I have described it, that very little evidence—in the form of superficially contradictory results—is required to indicate the presence of a hidden variable.

Thirdly, it is necessary to say something further about the "degeneracy," as Chomsky calls it, of the evidential data from which on an empiricist account a human child has to construct its internalized grammar. "The learner," writes Chomsky,¹³ "must select a hypothesis regarding the language to which he is exposed that rejects a good part of the data on which this hypothesis must rest." But the position here is precisely the same as in natural science. So called "fact-correcting" hypotheses are quite common at the level of generalization about correlations between natural variables. When such a hypothesis would have to suffer very many modifications and complications to retain full support from the evidence, we often avoid introducing these modifications and qualifications into the explicit formulation of the hypothesis by introducing them instead into our relatively tacit conception of the hypothesis' domain of discourse. Typically, a domain of individuals is postulated that are not subject to the operation of whatever variables would otherwise create exceptions to the generalization. A generalization about the acceleration of falling bodies, say, is conceived to apply only to objects in a frictionless medium. But what from the point of view of the real world is an idealization, because there is in fact no frictionless medium, must be seen, from the inductive point of view, as a restriction of meaning, because it is the price that has to be paid for freedom from a particular kind of unfavorable test-result. The domain of the generalization is a simpler one than really exists, but its characterization in familiar terms is

¹³ Noam Chomsky, *Language and Mind* (New York, 1968) p. 27.

correspondingly more complex since certain features of objects as we know them—certain effects they are liable to undergo—have to be excluded. Similarly an inductive language-learning device may be supposed to internalize a syntax that is true only of ideal speakers—speakers who are never affected by interruptions, changes of mind, shortness of time, limitations of memory, and other distractions. The syntactic competence acquired is undoubtedly one that must reject as deviant a good part of the data on which it rests. But a language-learning device may acquire such a competence empirically by procedures of qualification in the face of counter-examples that are quite straightforwardly inductive. The element of idealization that distinguishes the syntax of our language (according to most linguistic theories) from an exact representation of how we actually speak—i.e., that distinguishes competence from performance—is not an argument for the innateness of certain syntactic principles. This element of idealization is not at all a difficulty for an empiricist account of language-learning, as Chomsky argues, but just what one would expect to find if an empiricist account were true.¹⁴

Fourthly, it should be remembered that syntax-acquisition, like any other piece of human learning, will be more satisfactorily explained if it can be related to general learning strategies than if the innateness of some appropriately specific mechanisms has to be postulated. Here, as elsewhere in science, the more comprehensive pattern of explanation is the more desirable one. The presumption is that such an explanation is attainable, so long as we have no evidence or argument that unambiguously rebuts this presumption.

II

If the syntax of a language is to afford a subject for non-statistical generalization, it can certainly be

conveniently viewed as a system of rules rather than of facts, so that very many utterances may be written off as deviant or "degenerate" data. Yet most such utterances are intelligible, even if they sometimes seem ambiguous, and even the much more ungrammatical utterances of children and foreigners are very often fully intelligible. These facts present an obvious problem to any theory of speech comprehension that supposes the syntactic analysis of what is heard to be an essential stage in the process by which speech is understood. In particular, one might well expect an adequate theory of syntax to be able to embrace deviant strings by placing them on a scale of grammaticality that elucidates their relationship to fully grammatical sentences.¹⁵

I do not propose to discuss here the problem whether a speech comprehension device in fact requires to make a full grammatical analysis, or parsing, of its input in every case. I have suggested elsewhere that there is some psychological evidence that this does not always happen with human beings;¹⁶ and there is also some evidence of partial success with a mechanical translation program (at Seattle¹⁷) that dispenses altogether with syntactic analysis of source material. Indeed it is often the case that some of the syntactic constraints operating on a particular sentence are redundant for semantical purposes in that sentence, because more structuring is created than is necessary, in that sentence, for the requisite semantical load. It is not essential, e.g., to have a plural verb-form in *They eat here*. In such cases an attempt at a full syntactic analysis would have more point for the purpose of recovering the full text of a message that had been corrupted in transmission, or imperfectly stored in memory, than for the purpose of discovering its meaning.

I have shown elsewhere, however, that, if inductive reasoning is conceived as I have been suggesting, it can provide a framework for elucidat-

¹⁴ One sometimes hears it argued that there are certain rather recondite features of English syntax that (i) every adult English speaker knows despite the fact that (ii) many English speakers pass their childhood without experiencing any evidence of them. But, in default of empirical evidence in their favor, (i) and (ii) are speculations that afford no foundation for any argument. One has yet to read—in the psychological literature—of an English speaker for whom there is adequately attested evidence both that he knows some such syntactic feature and that he has never experienced any evidence for it.

¹⁵ Cf. Noam Chomsky, *Syntactic Structures* (Gravenhage, Mouton, 1957), pp. 35 f. and 78, and "Some Methodological Remarks on Generative Grammar," *Word*, vol. 27 (1961) p. 219 ff.; and J. J. Katz, "Semi-sentences," in *The Structure of Language*, ed. by J. A. Fodor and J. J. Katz (Englewood Cliff, N.J., 1964) p. 400 ff.

¹⁶ In J. Lyons and R. J. Wales (eds.), *Psycholinguistics Papers* (Chicago, 1966) p. 167 ff. Cf. M. Ross Quillian, "The Teachable Language Comprehender: A Simulation Program and Theory of Language," *Communications of the Association for Computing Machinery*, vol. 12 (1969) pp. 459–476.

¹⁷ Cf. E. Reifer, "Chinese-English Machine Translation: Its Lexicographic and Linguistic Problems," in *Machine Translation*, ed. by A. D. Booth (Amsterdam, 1967) p. 319 ff.

ing the concept of a scale of grammaticalness, on the lower ranges of which belong the syntactically deviant utterances that occur so commonly in everyday speech.¹⁸ Moreover within this framework it is possible to elucidate the precise nature of an ability that every normal adult speaker seems to have, viz., the ability to improve the syntax of a syntactically deviant sentence that he utters or hears uttered. It turns out that a language-speaker's ability to process semi-grammatical sentences, which any linguistic or psycholinguistic theory must take into account, can be seen as just one more manifestation of an intellectual ability that is also often manifested in other fields of inductive reasoning, wherever relevant evidence is available, viz., the ability to reduce a too simple hypothesis (which in this case upholds the semi-grammatical sentence) to a more complex but better supported version.

III

I come finally to the third of the three problems on which some light seems to be shed by the proposed account of inductive reasoning, viz., the problem of the origin, or explanation, of semantical categories. Certainly the constraints on sentence-construction that stem from differences between these categories are nowhere near as rigid or restrictive as those that stem from syntactic rules. Metaphors and conceptual innovations, unlike syntactic solecisms, have important and respectable tasks to perform in human communication. Nevertheless at any one time and place most nouns, verbs, adjectives, and adverbs have a central core of standard meaning or meanings by reference to which we judge a metaphor to be a metaphor or a conceptual innovation to be a conceptual innovation—i.e., by reference to which we judge certain collocations of words to be semantically abnormal. So here too a problem arises about language-learning. How can we explain the ordinary speaker's awareness of these semantical constraints on normal sentence construction?

Chomsky suggests that the explanation is at least in part to be found by supposing a considerable apparatus of innate semantical universals. "It is surely our ignorance of the relevant psychological and physiological facts," he writes,¹⁹ "that makes

possible the widely held belief that there is little or no *a priori* structure to the system of attainable concepts." What are his positive arguments for innateness? First, presumably, there is the old Chomsky-Miller argument about the time factor involved in inductive language-learning; and secondly, there is Chomsky's claim²⁰ that "the very notion 'lexical entry' presupposes some sort of fixed, universal vocabulary in terms of which these objects are characterized, just as the notion 'phonetic representation' presupposes some sort of universal phonetic theory." But both arguments can be shown to be invalid, in view of the role that can be played in language-learning by induction from variety of circumstance.

Note that an inductive language-learning device that is to afford a plausible model for children's behavior cannot be supposed to acquire a knowledge of semantical constraints in quite the same way as I suggested earlier that it might acquire a knowledge of syntactic (or morphological) ones. Children commonly utter deviant sentences in a way that may be construed as an experiment to determine the correctness of some set of syntactic or morphological rules. They say such things as **He singed loud* or **That's impossible*. But they rarely utter sentences that are just semantically abnormal in the course of ordinary speech, and if they do produce an occasional metaphor (e.g., *The car has gone to sleep in the garage*) it is more likely to be praised by adult hearers than corrected. It is certainly conceivable that in an inhumanly pedantic and prosaic speech-community an inductive learning-device could acquire knowledge of lexical constraints in quite the same way as I suggested for syntactic or morphological ones. But the device would then have learned to put metaphor on a par with grammatical solecism, which is a procedure that runs counter to normal views about the usefulness of metaphor and the uselessness of solecism.

How then can an inductive learning-device be supposed to acquire a knowledge of semantical categories and constraints? The answer to this question is not hard to find if we attend to the fact that a speaker is not just a sentence-producer but someone who utters sentences in appropriate contexts and for appropriate purposes. In particular he often names and describes things in or by his utterances, and so when he learns the use of nouns, verbs, adjectives, adverbs, etc., he is at the very

¹⁸ *The Implications of Induction*, (op. cit.), pp. 178-180.

¹⁹ Noam Chomsky, *Aspects of the Theory of Syntax* (Cambridge, 1965) p. 160.

²⁰ *Ibid.*

least learning how to name and describe things. In what way, then, can an inductive device be supposed to learn the English name, say, for motorcars? It is no use just attributing to the device appropriate sensory mechanisms and a fairly elaborate pattern-recognition component, on the assumption that it will then suffice for the one-word sentence *Motorcar!* to be uttered the next time a motorcar comes within the device's sensory field. For obviously the thing thus indicated might not be specifically a motorcar. It might be either an object or an event, and if an object it might be either any mobile object or only an inanimate one, and if an inanimate object it might be either any inanimate object on wheels or only a medium-sized one, and so on, just as if it were an event it might be either any movement or only self-propelled movement, and so on. But these philosophically notorious facts (cf. W. V. Quine on the problem of discovering the meaning of "gavagai" in an exotic language)²¹ need not constitute any stumbling block for a genuinely inductive learning-device. They merely reveal the relevant variables for testing hypotheses like those about what the word *motorcar* names.

Assume that an elementary hypothesis on this topic has initially the excessively general form:

Any conspicuously segregated object is named a motorcar.

Then again, just as with elementary syntax-learning hypotheses, there are two ways of proceeding. If all the relevant variables were known, and the possible combinations of their variants were not too numerous for the time available, the meaning-learner could discover a fully supported version of his hypothesis by carrying out the most thorough test of it and introducing modifications appropriate to his test-results. For example, if the relevant variables were plant/animal/artefact, aerial/aquatic/terrestrial, mobile/immobile, etc., etc., then the fully modified hypothesis would read

Any artefact, that is terrestrial, mobile, etc., etc., is named a motorcar.

But, as in the case of syntax-learning, there is also another way of proceeding, which may be more

suitable if many of the relevant variables are initially unknown and the time available is relatively short. This method consists essentially in conducting a series of separate tests (leading to appropriate modifications of the hypothesis) in relation to each relevant variable, instead of one big test in relation to all of them at once. What is vital in such a single variable test (which may itself be spread over quite a period) is that there should be some way of promoting an unambiguous interpretation for each unsuccessful trial-outcome. The learner needs to discover which relevant variable to blame for the falsification. So what he has to do is to perceive a difference between the conspicuously segregated object that he himself has wrongly named a motorcar and the one that he has heard others correctly so name. He may thus acquire both the concept of a particular variant of a relevant variable, if it has not been acquired already, and also learn how to modify one of his semantical hypotheses appropriately. Of course, the semantics-learner may be mistaken in what he takes to be an operative difference between the circumstances of his own, deviant utterance and the circumstances of some adult utterances. But if so the pressures of his culture can be expected to ensure that his mistake will emerge in a later trial-outcome.²²

It is now possible to characterize metaphorical usage in terms of the conceptual apparatus of inductive logic. To use a word metaphorically is to use it in accordance with a hypothesis that differs from the standard hypothesis for that word in respect to one or two variables but resembles it in respect to all the others. E.g., if a motorcar is said to have gone to sleep in the garage, then *sleep* here is being used metaphorically in respect to the plant/animal/artefact variable, though literally in respect to the active/quiescent variable. (The metaphorical meaning is intelligible, in the context of utterance, just so far as substitution of a literal usage in that context would have introduced an element of semantical redundancy.)

Note that a semantical hypothesis like

Any artefact that is terrestrial, mobile, etc., etc., is named a motorcar

does not assert a definitional synonym or create an

²¹ *Word and Object* (Cambridge, 1960) p. 51 ff. Quine seems to allow some element of variational induction in the determination of stimulus meaning (*ibid.* p. 40). But he seems not to see that the more thoroughly inductive investigators of an exotic language test their semantical hypotheses the less they have to admit indeterminacy in their translations.

²² Cf. for a similar thesis, L. S. Vygotsky, *Thought and Language*, tr. by Eugenia Haufmann and Gertrude Vakar (Cambridge, 1962) p. 56.

analytic truth. It does not need to. So long as it includes a variant of each of the variables that is in fact relevant to such a hypothesis, no inductive learning-device that has established the hypothesis will name anything a motorcar that is not one or be incapable of naming a thing a motorcar when occasion to do so arises. Perhaps someone will object that if by chance two nouns did apparently coincide in both extension and contextual suitability, though not in intention—i.e., in the class of individuals they were assumed to denote and in which kinds of social setting they were uttered, though not in their meanings—then an inductive learning-device could not distinguish the meaning or linguistic function of the one noun from the meaning or linguistic function of another. Quite so: it could not. But examples of such word-pairs are very hard to find. *Unicorn* and *centaur* are not a good example, because to say that both nouns denote the null-class and therefore coincide in extension is to ignore the fact that in these nouns' normal context of use, viz., in stories, myths, and fables, both unicorns and centaurs are assumed to exist. The normal examples of coincidence in extension but not in intention relate a word to a phrase (like *vertebrate* to *possessor of kidneys*), or a phrase to a phrase, rather than a word to a word. If there are pairs of nouns (or pairs of adjectives, or pairs of verbs, etc.) in any language that bear this relation to one another, I do not see how their meanings can ever be taught or learned unless at least one member of the pair is introduced not by a process of inductive reasoning, but implicitly or explicitly, as a synonym for a phrase containing two or more other words.

This is not the place to attempt any solution of the very large problem of precisely how the semantical normality or abnormality of whole sentences may be judged. But it is clear that, whatever else any such solution requires, it cannot get off the ground without assuming some fairly detailed semantical characterizations or categorizations of lexical items, and my point here is that elementary, inductively acquired hypotheses about how to name or describe things can supply such characterizations. To mention a variant of a variable that is inductively relevant to such a hypothesis is to cite a semantical marker, and a hypothesis of this type that has been sufficiently modified to enjoy full support supplies a set of semantical markers that can normally be regarded

as uniquely identifying the central meaning or meanings of the noun, verb, adjective, or adverb in question. Indeed such a set of markers is indistinguishable, in its potential as a basis for judgments about the semantical normality of sentences, from a distinctive feature matrix of the type that Chomsky proposes.²³ The distinctive features of phonology, say, or lexicology are the (two-variant) variables that are supposed to be inductively relevant in those fields. What I have been trying to show, in effect, is that an adequate account of inductive reasoning can provide an explanation for the validity of such distinctive feature matrices that does not need to invoke the supposition of innate linguistic universals. But, again, as in the case of syntax-learning, I have not been claiming that this is in fact how in practice children can or do learn the meanings of their native words, or linguists learn the meanings of words in an exotic language. I have been claiming only that a certain type of argument for innate linguistic universals is invalidated.

No doubt borderline cases can in principle arise in relation to any such hypothesis about how to name or describe things: are stools, for example, to be regarded as a backless form of chair or as not being chairs at all? But this well-known phenomenon, which Waismann called "open texture,"²⁴ confronts any semantic theory, and though it creates substantial difficulties for any attempts to take the semantics of formalized languages as a model for the semantics of natural language it creates no difficulty at all for the present model of semantic learning. A borderline case is to be viewed as a challenge, or potential counter-instance, to a hypothesis that was thought to be fully supported, and as such it reveals the possibility of a so-called "hidden variable," a variable of hitherto undetected relevance. For in determining whether the borderline object or situation should or should not be taken to fall within the denotation of the lexical item in question we obviously have to decide whether some hitherto unregarded consideration is relevant to the use of that item—e.g., whether having a back or not is relevant to whether a thing should be described as a "chair."

It may be of interest to note that the proposed conception of semantical categories, as the relevant variables for inductive reasoning about how to name or describe things, has another advantage too. It contrives to reconcile the linguist's natural

²³ *Aspects of the Theory of Syntax*, op. cit., p. 148 ff.

²⁴ F. Waismann, "Verifiability," *Proceedings of the Aristotelian Society*, Supplementary Volume XIX (1961) p. 236 ff.

desire to arrange these categories in a hierarchy of importance with the implausibility of such an arrangement by progressive subdivision of higher categories. Chomsky originally proposed, as a basis for assessing degree of grammaticalness, a single, many-leveled hierarchy of sub-categorizations, that would gradually split down the most general categories (e.g., noun) into the most specific (e.g., noun, animate, human, female, young).²⁵ But, as Chomsky himself soon came to see,²⁶ the frequent need for cross-classifications makes such a hierarchy of sub-categorization impossible.

Once semantical categories come to be viewed as inductive variables, it immediately becomes apparent how it is possible that, on the one hand, the categories belong in a hierarchy of importance, and, on the other, the features that belong to different categories can occur in all sorts of combinations with one another. The hierarchy of importance is the order of inductive relevance, the like of which we can find in any field of inductive reasoning whatever: in experimental science, e.g., as well as in language-learning. The multiform combinations—or distinctive feature matrices, as Chomsky calls them—are just the combinations of variants of different variables that constitute the circumstances in which the most simple and unqualified semantical hypotheses admit of being tested. Here as elsewhere in inductive reasoning the most thorough test of a hypothesis requires it to be tried out over all possible combinations of variants of relevant variables, while a typical semantical hypothesis that can enter into what Chomsky calls a “lexical entry” is a heavily qualified hypothesis that has been guaranteed full inductive support by the insertion into its antecedent

of an appropriate combination of variants of relevant variables.

Moreover, so far as there is a “fixed, universal vocabulary,” as Chomsky calls it, in terms of which lexical entries need to be formulated, its existence is to be accounted for by the fact that the relevant variables for hypotheses about how to name or describe are sets of features of reality, not of language: e.g., object/event; animate/inanimate; natural/artificial; etc. So far as speakers share a common non-Whorfian world, we must expect to find a common vocabulary appropriate for the formulation of lexical entries in the description of their languages, and we do not need to postulate innate linguistic universals to account for the appropriateness of this common vocabulary. What has perhaps sometimes tended to obstruct appreciation of that fact is the rather off-center conception of a language speaker’s competence as the ability to generate grammatical sentences and map readings on to them.²⁷ When instead we view this competence as including—alongside the generation of grammatical sentences—the ability to judge the appropriateness of sentence-utterances to their socio-physical settings, and so as including the ability to name, describe, refer,²⁸ etc., etc., we can see more clearly how the semantical structure of language is rooted in the structure of the world rather than of the human brain. Or at any rate, so far as it does depend on the structure of the human brain, it depends on the ways in which we can perceive and control the world, not on specifically linguistic innate principles. So here too an argument of Chomsky’s for specifically linguistic innateness seems to be rebuttable by a proper appreciation of how it is possible for inductive reasoning to operate.

The Queen’s College, Oxford

Received December 26, 1969

²⁵ “Some Methodological Remarks on Generative Grammar,” *Word*, vol. 17 (1961) p. 119 ff.

²⁶ *Aspects of the Theory of Syntax*, *op. cit.*, p. 79 ff. Cf. L. Jonathan Cohen, *The Diversity of Meaning*, 2nd ed. (New York, 1966) p. 78.

²⁷ E.g., Noam Chomsky, *Topics in the Theory of Generative Grammar* (The Hague, 1966) p. 10. It is doubtful whether in fact a normal speaker has not the ability to map readings onto the sentences of his language, in the sense in which he obviously has got the ability to generate grammatical sentences. The ability to paraphrase explicitly is a skill sometimes taught in schools, but many find it very difficult to practice. Nor does it avail to claim that a hearer can understand what is said to him only by making a tacit paraphrase, since an infinite regress then arises if we ask how the hearer understands his own paraphrase. Moreover conceivably, if we knew enough about English grammar, we could teach a man to generate English sentences and to map English readings onto them without his ever being able to think or communicate in English at all. So having the ability to generate English sentences and paraphrase them is neither a necessary nor a sufficient condition for being an English speaker.

²⁸ On the ability to refer cf. L. Jonathan Cohen, “What is the ability to refer to things, as a constituent of a language-speaker’s competence?” in *Languages in Society and the Technical World* (Proceedings of the Olivetti Centenary Symposium at Milan: to be published).

IV. PURPOSIVE ACTION

BERNARD BEROFISKY

THERE are two radically different theories as to the nature of purposive action, i.e., as to the analysis of "*P* does *A* in order that (for purpose) *x*." The causal thesis, as I shall call it, maintains that the analysis will take the form of a statement asserting a causal relation between psychological states of *P* and his doing of *A*. Hence, the explanation of the action provided by the reference to *P*'s purpose is of the same type as the explanation of a machine's behavior, i.e., causation by internal states. A person holding the causal thesis may, of course, believe that man's internal states are distinctive, e.g., that only man has desires and beliefs.

The other theory, which I shall call primitivism, maintains that the explanatory relation between an act and a purpose is not reducible to a causal relation but is rather distinctive. Thus, man is distinctive both in his possession of desires and beliefs and in the way these desires and beliefs account for his actions.

Ducasse has defended the causal thesis. His analysis of "*P* does *A* for purpose *x*" is roughly "*P* does *A* and his doing of *A* is caused by a desire for *x* and a belief that *A* will achieve *x*."¹

Chisholm has argued for primitivism.² Insisting that explanation of purposive behavior is *sui generis*, he introduces as a *primitive* notion the idea of doing something in the endeavor to do something else and defines a number of concepts in its terms: doing something for a certain purpose, the notion of a successful intentional action, etc. This set of concepts cannot be reduced in the way Ducasse advocates, i.e., to causal relations between states and behavior.

An important advantage of the Ducasse view is that his analysis uses a familiar concept, viz., causation. Thus, the puzzling notion of purposive action is explained in terms of a concept we understand or, at least, have to try to understand anyway. But Chisholm refuses to allow that the family

of concepts he is interested in can be clarified further. In this way, he establishes the uniqueness of psychological explanation—where purposive behavior is concerned, one is on a radically independent plain of concepts.

In this paper I shall defend the causal thesis by showing that *prima facie* counterexamples to a version of it can be incorporated into modified versions of the causal thesis. In the course of the discussion, I shall also reply to a serious charge of circularity. The only other arguments against the causal thesis with which I am familiar are designed to show that the claim that a causal relation between desires (beliefs) and actions exists is incompatible with the fact that certain logical or conceptual relations hold between desires (beliefs) and actions. I have tried to show elsewhere that these arguments all fail,³ and if that is right, this paper should suffice to establish the causal thesis. At least, the burden of proof falls upon primitivism.

Since the causal thesis is a claim about the form any adequate analysis of purposive action must take, I shall produce an actual analysis and see whether or not it is an extensionally adequate one. The analysis I shall defend differs in one significant way from Ducasse's version. Chisholm points out that a man may set out or endeavor to do what he does not want to do. So a man can do something for a certain purpose even if no desire is motivating him. But Ducasse's analysis requires that a purposive act be caused by a desire.

I think Chisholm is right. Normally, when a man does something in order to effect some end, we suppose he wants that end. Even if it turns out that he is doing something he really does not want to do, we can find some *reason*, e.g., he feels obliged to do it, or he is doing it to prevent some greater evil, e.g., going to the dentist. But there is no logical absurdity in supposing that a man does something in order for some end to come about without having *any reason*. Suppose I reach out to pick up a

¹ C. J. Ducasse, "Explanation, Mechanism, and Teleology," *Readings in Philosophical Analysis*, ed. by H. Feigl and W. Sellars (New York, 1949), p. 543.

² R. M. Chisholm, "Freedom and Action," *Freedom and Determinism*, ed. by K. Lehrer (New York, 1966), pp. 11-44.

³ *Determinism* (Princeton University Press, forthcoming).

piece of dust. Does it *have* to be the case that I wanted to do that? Maybe I did, but maybe not. Surely there might be no conscious desire. There might have been an unconscious desire; but this requires substantiation; therefore, the attempt to substantiate it might fail.

Some act descriptions, of course, logically imply the existence and explanatory efficacy of a reason, e.g., "He is satisfying his sexual desire." But even if we suppose that this act cannot be redescribed without any implications regarding its explanation, as some theocrats believe, I am simply saying that *some* purposive actions require no reason.

What I think has happened is this: When a man does something for some purpose, the explanation we seek is in terms of the man's desires or reasons. And we suppose that if we dig deeply enough, some reason will appear. Now this suspicion may be correct. But we ought to distinguish the *nature* of purposive action from the normal *explanation* of purposive action. I shall, therefore, drop the concept of desire from my version. Let me present an analysis that has some initial plausibility even if it turns out to be ultimately unsuccessful, and then explain some key terms in it:

$$P \text{ does } A \text{ at } t \left\{ \begin{array}{l} \text{in order that} \\ \text{for purpose} \\ \text{with the intention that} \end{array} \right\} x = \text{Df}$$

1. P does A at t

2. P believes at t that A has some chance of bringing x about

3. (2) is a necessary condition in *esse* of (1)

1. "Bring about" can mean "lead to" or "satisfy." In other words, the definition is supposed to cover cases in which a man does something so that it will lead in the future to some consequence, e.g., I invest in the stock market to get rich, and cases in which the purpose is the act itself under a different description, e.g., I give you this money in order to pay back a debt.

2. (2) is often expressed more strongly, e.g., P believes that A will probably bring x about. But it is clear that a man might act purposively where he has only a slim chance of success. A man might undergo an operation to save his life even if he has say a 5 per cent chance of success if his chances of survival are zero without the operation.

3. Analyses, like Ducasse's, very often use the term "cause," but do not submit it to any more careful scrutiny. Instead of "cause," I am using the

more specific concept of a necessary condition in *esse*.

c is a necessary condition *in esse* of e if and only if c is a necessary condition of e and c exists at every moment e exists. Thus, at any moment during e 's existence, e would cease if c were to cease. Being born is a necessary condition of a person's being five years old, but not in *esse*. Oxygen is a necessary condition in *esse* of fire because fire would cease whenever oxygen is removed. Oxygen, in other words, sustains fire, while being born does not sustain being five years old.

c is a necessary condition of e in a situation S if and only if S is an instance of a general (non-logical) law that entails the occurrence of c given e and a set of conditions f in S . Since action-types can be performed for an indefinitely large number of purposes, the belief-token described in (2) is necessary to the action-token described in (1). Hence, the applicable general law takes the form "(1) and f implies (2)" (where " f " is instantiated in the situation) rather than "(1) implies (2)."

We may also stipulate that where the analysis fails for some belief b_1 , but succeeds for a disjunction of beliefs of the form of (2) that includes b_1 (where each belief in the disjunction is non-vacuous in the sense that its omission from the disjunction makes the disjunction fail the analysis), then the action is multipurposed. Thus, if I invite friends to dinner in order to satisfy an obligation and in order to have a pleasant evening, it may be that neither purpose is necessary to the act of inviting them in that situation, but the disjunction of purposes is.

It is to be noticed that the concept of a necessary condition in *esse* is applicable to inanimate physical nature as well as human nature. Thus, the only conception in the definition that has unique applicability to persons is the concept of belief. We have dropped the concept of desire, and we have preserved the idea that no special relation between action and its explanation exists in the human sphere. Thus, if purposive action is man's distinctive trait, on my view, that turns out to mean that man is the only creature with beliefs and that these beliefs can be efficacious in the dull and universal sense of causal efficacy, or, more specifically, the efficacy of necessary conditions in *esse*.

The reason that the belief is a condition *in esse* will become clear when we consider the objections to the extensional adequacy of the analysis.

Objection 1: (R. Taylor⁴) A man in an audience

⁴ R. Taylor, *Action and Purpose* (Englewood Cliffs, N.J., 1966), p. 249.

wants to attract the speaker. This state makes him fidgety and he thereby attracts the speaker. But he did not fidget *in order* to attract the speaker. Moreover, he may believe (correctly) that his fidgeting would attract the speaker.

My analysis deals successfully with this case. It is true that the man did not fidget in order to attract the speaker because, although he believed that fidgeting would attract the speaker, this belief was not a condition of the fidgeting. If his desire to attract the speaker made him fidgety, the belief that he would attract the speaker was irrelevant and, if anything, might have inhibited the behavior. But if it is true that he would have ceased fidgeting upon coming to believe that the speaker is not being attracted by it, then he really is fidgeting in order to attract the speaker.

But, it may be objected, there are cases in which a man who wants very much not to attract the attention of the speaker fidgets ("involuntarily") *because* he believes that fidgeting would attract the attention of the speaker. The belief, in other words, makes him nervous and causes the fidgeting; but he is surely not fidgeting in order to attract the speaker.

It seems clear that this man did not fidget in order to attract the speaker because he did not fidget intentionally. A similar case would be that of a pianist who believes that he will play badly if he looks at the music and looks at the music (unintentionally) just because he has this belief.⁵ Since a man does not do *A* in order that *x* unless he does *A* intentionally, we ought to modify (1) in the analysis to read "*P* does *A* intentionally at *t*."

We now face the dual task of analyzing the appropriate sense of intentionality and replying to the following charge of circularity.⁶ Even if some causal analysis is extensionally adequate, it will not succeed in reducing purposiveness to a causal relation between states and behavior because the states, viz., beliefs or beliefs and desires, must be construed as causes of actions and the concept of action here is intrinsically purposive or intentional.

There are, it seems to me, at least two concepts of action, one purposive, the other non-purposive. I may wish to distinguish the fact that I raised my arm, where this act may have been performed unintentionally, unthinkingly, out of force of habit, etc., from the fact that my arm just rose. That is, we may wish to distinguish doing from undergoing in

a sense whose elucidation requires no appeal to the concept of purpose. I have nothing illuminating to say about this sense of "action." But since it is a legitimate sense and since it involves no covert reference to purpose, I may use it. Call this "action₁."

We may now use the causal analysis to define the purposive sense of action: "I raise my arm intentionally (on purpose)" means "I raise my arm (action₁ sense) and this action has as a necessary condition in esse the belief that I raise my arm." In other words, to do *P* intentionally is, roughly, to do *P* "because" you believe you are doing *P*.

But there is a difference between doing *P* intentionally and having *P* as one's intention. It is not my intention to wear down the carpet when I walk to the other side of the room to get a book. But I knew that I was wearing down the carpet and I could have walked around it. I can be held responsible by a neurotically fussy housekeeper. Hence, there is a sense in which I wore the carpet down intentionally although wearing the carpet down was not my intention. Since I would have worn the carpet down even if I had come to believe that I was not wearing it down (I am misinformed about the carpet's properties), this case fails my analysis. Hence my analysis analyzes the stronger claim that such-and-such is the man's intention. In other words, "Jones raises his arm and this action has as a necessary condition in esse Jones's belief that he raises his arm" analyzes "In raising his arm, it is Jones's intention to raise his arm."

That is all well and good because the stronger claim is the one implicit in purposive action. If my belief that I was wearing down the carpet was causally irrelevant to the act, I could not have been wearing down the carpet in order, say, to anger the housekeeper.⁷ Hence, the concept of action implicit in purposive contexts can be analyzed in terms of a causal relation between two beliefs, a belief about what one is doing and a belief about the outcome of what one is doing, and the action itself, the latter being understood in a non-purposive sense.

The weak sense of intention whereby I wore the carpet down intentionally may be understood simply in terms of the presence of the belief (knowledge) that I was wearing down the carpet plus, perhaps, the fact that I could have acted otherwise.

⁵ I am grateful to R. Shope for this example.

⁶ The charge was leveled by A. Collins in conversation.

⁷ If the modification mentioned under objection 2 is adopted, this statement will have to be amended accordingly.

It is well known that the latter requires clarification; but we may need it here if doing something intentionally (even in the weak sense) implies responsibility.

This analysis of intentionality enables us to say that the man who fidgeted involuntarily did not act intentionally because the belief that he is fidgeting was not a condition of the fidgeting. But it is not clear that all such cases are so easily handled. May it not be the case that the pianist would not look at the music he is playing unless he believes he is looking at the music he is playing? Or can a neurotic husband claim that he does not intentionally insult his wife although he would not do it unless he took himself to be insulting his wife?

These two cases differ in important respects. The neurotic husband's claim is plausible because, as we suggested earlier, intentionality suggests the power to act otherwise and the neurotic husband presumably lacks that. Even so, he may be insulting his wife in order to get her goat, and we said earlier that a man does not do *A* in order that *x* unless he does *A* intentionally, in which case it follows that the husband's insults are intentional. This result is not really counterintuitive were it not for the suggestion of the freedom to act otherwise. (A psychoanalyst would certainly say and the husband might come to agree that his insulting is intentional.) Since purposiveness is our primary concern, and since his behavior is clearly purposive, we shall say that his insults are intentional (for this result is not, as we have seen, fundamentally counterintuitive).

The other case includes an interesting feature that seems to account for our regarding the act of looking at the music as nonintentional. The act takes place only because the person wants it not to (although he is not acting out of a general resolution to do what he does not desire to do). Although I have argued that a man may do *A* intentionally even though he does not want to do *A*, he does not do *A* intentionally if he does *A* out of the very fear of doing *A*—unless another condition of *A* is his general resolve to do what he fears (or the specific resolve to do what he fears in this case).

Thus, there are connections of a sort between desire and intentionality and we shall note others later. I only object to the claim that intentionality or purpose implies desire.

Once these modifications are incorporated into our analysis of intentionality, it appears that this notion can be analyzed without recourse to any of Chisholm's primitive ideas.

⁸ *Op. cit.*, pp. 29–30.

Objection 2: (R. Chisholm⁸) A young man wants the money he would inherit if his uncle were to die and believes he would get the money if he were to kill his uncle. This desire and belief agitate him so severely that he gets into his car, drives recklessly, and accidentally kills a man he later learns is his uncle. So the desire for money plus the belief that killing his uncle would satisfy the desire causes the young man to kill his uncle. But obviously he did not kill his uncle *in order to* get the money. This interesting case causes us no difficulty because the man did not kill his uncle intentionally. Moreover, condition (2) requires that there be some belief *at the time the act is performed* of the form "this act has some chance of getting me my uncle's money." If the act is described as "running over a man," the nephew does not believe that this act will make him rich because he does not know that the man is his uncle. If the act is described as "running over his uncle," then at the time of the action, the man does not believe that the act of running over his uncle will make him rich because he does not believe that he is running over his uncle. He has the general belief that if he ever kills his uncle he will be rich. But my analysis requires that he have a specific belief about the act he is performing and a man cannot believe that his specific action *A* will lead to *B* if he does not believe he is doing *A*.

I am not maintaining the patently false position that a man who is doing *A* must believe he is doing *A*. I may sit down on a wet seat without realizing it is wet. But I cannot sit down on a wet seat in order to have a good excuse for changing my slacks if I do not believe I am sitting down on a wet seat.

But what about a man who turns on his radio in order to hear music, oblivious of the fact that the radio contains tubes that must warm up if the radio is to work. We may say that he is heating up tubes without realizing it when he turns on the radio. May we also say that he is heating up tubes in order to hear music? If so, a man may do *A* in order that *x* without believing he is doing *A*.

If we accept this counterexample, we can revise the analysis by converting (2) into a disjunction of the original (2) and: There is an *A'* such that *P* does *A'*, *A'* is causally relevant (necessary or sufficient) to *A* or in fact identical with *A*, and *P* believes at the time he does *A'* that *A'* has some chance of bringing *x* about. We shall then have to revise (3) by adding the disjunct: *P*'s belief at the time he does *A'* that *A'* has some chance of bringing *x* about is a necessary condition in esse of (1).

This revision allows us to say that a man heats up tubes in order to hear music, but does not require us to say that the nephew killed his uncle in order to inherit the money. If *A* is "running over a man," the nephew neither believes that running over a man has some chance of getting him his uncle's money nor does he believe at the time that he is doing something related in the specified way to *A* that may result in the desired consequence (although he is). If *A* is "running over his uncle," since he does not believe he is doing this, he does not believe that running over his uncle has some chance of getting him the money. Nor is there an appropriately related act that he believes will have this result (although there is).

Shall we accept this revision? Since I am really concerned about the causal thesis rather than my particular version of it, this question does not interest me. I have established that the revision would be in line with the causal thesis. Hence, the causal thesis is preserved regardless of whether or not we wish to suppose that the man warmed up the tubes in order to listen to music. There will be analogous occasions in this paper where I shall try to demonstrate that *prima facie* counterexamples to my original version of the causal thesis can be considered genuine ones to that version, but not to the causal thesis as such. Let us now introduce the term "emasculatation" for any demonstration of the impotence of an objection vis-a-vis the causal thesis as such.

(If we accept this revision, we shall have to allow that a man can do *A* in order to *x* where he does not do *A* intentionally, although there always will be an action *A'* related in the specified way to *A* that he does intentionally.)

Even if we waive the above considerations and allow that some appropriate belief was a condition of the man's act, the case fails (3) because the man's belief that killing his uncle would make him rich was not a condition in esse. Cases like Chisholm's where beliefs and desires set a chain of events going are advanced by others against Ducasse-type analyses. But these criticisms fail to recognize the way beliefs and desires function in purposive behavior. Beliefs and desires do not only initiate action; they sustain action and it is in their sustaining role that they are purposive. If I have a sudden urge to do something and do it an hour later I will not be doing it in order to satisfy the urge if I no longer have the desire at the time the action takes place. Or if my belief that Gold Eagle will win the

race causes me to go to a certain window at the track where I meet a friend who convinces me that Gold Eagle will lose and that it is good to bet on a loser because the track needs the money, then my belief that Gold Eagle will win the race led to my betting on him. But I am not betting on Gold Eagle in order to win some money since I no longer have this belief at the time of action. I have tried to incorporate this sustaining role of beliefs in the analysis by requiring that the belief be necessary at the time of the act, i.e., the act would cease whenever the belief ceases.

Suppose I believe Gold Eagle will win and am about to plunk down \$2.00. All of a sudden I recall that Gold Eagle is a mudder and it is a sunny day. There are several possibilities: (1) I may take back my \$2.00. So the belief that Gold Eagle will win turned out to be a condition in esse of the (hypothetical) act of betting \$2.00 for the act ceased when the belief ceased. And we would say "I was about to bet on Gold Eagle in order to win money." (2) I may, for a number of reasons, leave the \$2.00. I am embarrassed or it is too troublesome to go to a different window, etc. If asked "Why did you beg \$2.00 on Gold Eagle?" it would be clearly wrong for me to say "Because I thought he would win." The belief that he would win was only a condition in esse during the early stages of the action. The act had a momentum of its own; so that I did not really bet on Gold Eagle because I thought he would win. The truth is that I bet on Gold Eagle because I was on that line and I became embarrassed or I was lazy, etc. Moreover I was on that line initially in order to bet on Gold Eagle in order to win some money. Thus, so long as the belief that Gold Eagle will win sustains my act, I am doing whatever I am doing in order to win money. Should the belief change, the explanation of my action would have to change. Or suppose a man jumps from the Brooklyn Bridge in order to commit suicide. He may, as he falls, come to believe that he will not be killed; but he does not, therefore, return to the bridge. Here, too, we can say that he jumped in order to commit suicide. But the explanation of his fall is provided by physics if we want to know why unsupported bodies fall or by the fact that he jumped if we want to know what initiated this particular fall. In either case, he is not falling in order to commit suicide.

(Some act-descriptions apply after we cease to have direct control over them, e.g., he is hitting the target with his arrow.⁹ In these cases, a man's

⁹ I am indebted to R. Shope for this example.

beliefs are pertinent only up to a certain point. Evidently some appropriate restriction can deal with these, e.g., our thesis does not apply to those actions or parts of actions that a man can perform after he has died.)

Objection 3: There are many acts I would not perform if I discovered that they would have undesirable consequences. For example, I would not open the refrigerator if I thought it would blow up. On the proposed analysis, therefore, I open the refrigerator so that it does not blow up, rather than so that I can get an apple.

This criticism fails to notice that I must believe that my act will have a certain consequence if the consequence is its purpose. The failure of the refrigerator to blow up is not a consequence of opening it. The refrigerator did not blow up when I opened it; but this does not mean that I caused it not to blow up. There are many states the world is in at the time the refrigerator is opened, negative and positive, that I did not cause. Of course, my purpose might have been to *prevent* the refrigerator from blowing up, in which case, as the analysis says, I would not have opened it had I not believed there is a chance I would thereby prevent it from blowing up.

Objection 4: (R. Abelson¹⁰) A man may skydive in order to act courageously, i.e., in order to do something life-threatening. So when the man jumps, he believes there is a chance he will get killed and he would not jump if he did not have this belief. For if he thought there was *no* chance he would be killed, the act would lose its interest—he would not be displaying great courage. On my analysis, therefore, we would be forced to say that he skydives in order to be killed and this is clearly not so.

This objection can be answered by adding a plausible condition. Because of cases like the man who undergoes an operation that has only a 5 per cent chance of success, we were forced to weaken condition #2 so that a man must only believe that the act have some chance of success. But clearly if a man is doing *A* in order to *x*, he will not be averse to doing *A* if he learns that *x* will definitely result from his doing *A*. Hence a fourth condition is plausible: If *P* were to believe that *A* insures *x*, he would still do *x*. The skydiving case fails this condition. If the skydiver learns that his death is insured by jumping, he will not jump. Or if he does, then he really is jumping in order to commit sui-

cide. In the case where he does not want to commit suicide, he is jumping in order to do something life-threatening because he believes that the act will *insure* that he does something that has a small chance of taking his life.

Objection 5: (James Rachels¹¹) Smith, a professional gambler who loves his style of life, will not place a bet on a really sure thing since the thrill of gambling would not be present. Hence, because of condition (4)—just added to deal with the fourth objection—Smith does not bet in order to make money. But this consequence is odd given that Smith would be very disappointed if he were to lose.

Smith's purpose is complex: he wishes to win money in a situation in which he might lose. And this is his purpose on my analysis. He is different from Jones who would bet on a sure thing and my analysis describes this difference by saying that Jones bets in order to win money; Smith does not, but Smith is betting to bring about the complex result "uncertain victory." If, on the other hand, Smith bets when he knows he will lose, then he is obviously not betting to win. (Thus, Smith's belief would be expressed conjunctively whereas a dual purpose would be expressed disjunctively.)

Objection 6: Condition (2) is too strong because a man may act for an end even when he knows he will fail. A person in a concentration camp, for example, may feel that he must try to escape even if he is convinced that his attempt is futile.

In order for this case to be a counterexample, the man must really believe at the time of action that he has no chance of success. A man may believe this before the action, but not as he really proceeds with the attempt. Secondly, we must distinguish the man whose purpose is escape* from cases of "going through the motions," i.e., cases of people whose purpose is the expression of their dignity or a certain effect on other inmates or, perhaps, suicide.

It is not easy to decide whether or not there can be a genuine counter-example, therefore. This man must be bobbing and weaving in order to escape, and also be absolutely convinced that his efforts are futile. But this objection, I believe, can be emasculated.

For if a person believes that this man is bobbing and weaving in order to escape, he will change condition (2) to: Either *P* believes at *t* that *A* has some chance of bringing *x* about or *P* believes at *t* that *A* would have had some chance of bringing *x*

¹⁰ In conversation.

¹¹ In conversation.

about under different conditions. This new causal thesis is only slightly weakened because, for example, a man going to college in order to become a lawyer is not also going in order to become a doctor just because he believes that, had he wanted to be a doctor, going to college would probably lead to his becoming a doctor. This belief must also be a condition of his going to college (in order to become a lawyer) and it is not.

Objection 7: (R. Abelson¹²) We say of animals that they act for ends; yet we do not ascribe beliefs to animals, and, *a fortiori*, we do not ascribe beliefs regarding probabilities.

But we do, in everyday talk, ascribe beliefs to animals. We hesitate, upon reflection, to think of these beliefs as full-blooded because they do not involve dispositions to talk, i.e., to use language. But dispositions to act and, to some extent, to feel are present and so animals have beliefs in an emasculated sense of the term. The absence of language restricts the possible content of an animal's belief and makes us, therefore, hesitate to ascribe beliefs regarding probabilities to animals. Hence, the analysis of "in order to" talk for animals will have to refer to beliefs of a simpler structure.

I do, however, think that ascription of purpose and ascription of belief go hand in hand. If we became convinced that animals cannot have beliefs in any sense, we would construe ascription of purposes to them as anthropomorphic. In this regard, it is interesting to note that, according to C. Taylor, our ascription of purposes to animals is justified by the failure of psychologists to develop adequate theories that eschew intentional concepts like belief.¹³

Objection 8: There are several serious difficulties that fall under a single heading we might label the problem of subsidiary purposes. Suppose I have worked out in detail an itinerary for a trip from New York to California and, as I start out, someone asks me why I am going into the Lincoln Tunnel. I might specify an immediate purpose, e.g., to get to the New Jersey Turnpike, or I might specify the ultimate goal, to get to California. But it would sound very odd to say "to get to the Indiana Turnpike." On my analysis, however, all of these answers are equally legitimate because, on my analysis, they are all my purposes.

Now this difficulty is not terribly serious. There is a sense in which my purpose is to get to the

Indiana Turnpike. The reason it sounds odd is that the selection of one of so many intermediate goals seems arbitrary. So we do use criteria to select certain purposes and rule out others and my analysis has not taken this into account. Why?

There are three reasons. (1) As I just said, there is a sense in which every intermediate step necessary to the attainment of a goal is a purpose in some sense. (2) There is an element of personal relativity that plays an important role in the selection of the purpose or purposes. The selection varies with the way in which an individual classifies or divides his situation. If, on my itinerary, I thought of the Indiana Turnpike as the end of the first leg of the journey, it would not be unnatural to speak of arriving there as my purpose. Or the division of the activity may have a lot to do with social or legal conventions. For example, if I were asked my purpose in taking the Saw Mill River Parkway this afternoon, I could say "It will get me to the Henry Hudson" or "It will get me to the West Side Highway" or "It will get me to New York"; but it would sound very odd to say "It will get me to the section of the Saw Mill that goes through the Northern half of Yonkers" because that section of road has no special name. There is also a contextual component in our selection of the end or final goal of our activity. On my analysis, if I would have given up the trip to the Indiana Turnpike upon learning that it will not get me to California, then getting to California is a goal; but if I would have given up the trip to California upon coming to discover that California would not make me happy, then happiness is also a purpose of my entering the Lincoln Tunnel. So it looks as if Indiana and California are on a par. But there is still a sense in which the stages of the trip are merely means to the *goal* of getting to California. (3) Other principles are at work; but none of them is sacrosanct in the sense that the personal and contextual considerations mentioned above can override the failure of these principles as the following examples show.

One principle seems to be that a consequence to which the person has an aversion, but which he believes to be a means to something he wants, will not be called a purpose. I do not enter the store to rid myself of some money; but that is the only way I can get what I want. Here, again, is a connection between purposiveness and desire. If, however, the

¹² In conversation.

¹³ C. Taylor, *The Explanation of Behavior* (London, 1964), ch. 3.

necessary evil requires effort or concentration or if I dwell upon it for certain reasons, it may become a purpose of mine. A person who works at an unpleasant job only to earn money certainly sees himself as leaving the house in order to go to work. Consider another principle. We do not normally view a consequence of our action that is believed to be necessary for some goal as a purpose if the consequence is a natural, unlearned by-product of our activity. I may believe that I would not eat an apple unless I first salivate; but I do not think of myself as going to the refrigerator in order to salivate. Similarly, however, we can imagine a case where salivation is a goal, e.g., somebody asks me to salivate and I can only do this at the sight of food. So I go to the refrigerator to get an apple. I may in fact have a dual purpose if the request was made just before I had decided to get an apple from the refrigerator in order to eat it.

Evidently there is room here for work. It is clear to me that I have only scratched the surface in regard to this problem of subsidiary purposes. But, and this is the important point, I cannot believe that subsequent investigations will necessitate the abandonment of my analysis of purposive action. They will require modification and refinement of that analysis. The case for a Chisholm-type analysis, however, it seems plain, must be made on other grounds. And we have examined these grounds already and found them wanting. I would claim, therefore, that Objection 8 has been emasculated.

Objection 9: (M. Slote and E. Leites¹⁴) Suppose a man believes that if he were to write a certain sort of letter to Richard Nixon, Nixon would probably proceed to escalate the Vietnamese war. He does not write this letter for some reason whose nature is not at the moment pertinent. The most obvious reason would be that he does not want the war to be escalated. He is then hypnotized and the hypnotist commands him to do whatever he believes will probably escalate the war. When he wakes up, he proceeds to write this letter. Call this Case I.

On my analysis, we would have to say that he wrote the letter in order to escalate the war even in the case where he wants the war to end.

An even stranger case would be one in which the man does not believe the letter will have the escalating effect; is caused by the hypnotist to have this belief, and again is commanded to do whatever he believes will probably produce an escalation of the war. Call this Case II.

On my analysis, we would again have to say that he wrote the letter in order to escalate the war even if he wants the war to end.

I call these cases "alleged counterexamples" because it is not clearly false that the letters are written in order to escalate the war. The belief that the letters will have this effect must be included in an explanation of the action. It is not enough to cite the fact that the hypnotist commanded the man, say Jones, to do whatever he, Jones, believes will escalate the war. To explain the act of writing the letter, we must mention the fact that Jones believes that this specific act will probably bring on escalation. In Case II it is not enough to say that the hypnotist caused him to have the belief unless you suppose, per my analysis, that the belief is guiding the action in the sense that it is still present when he acts and is a condition of the act. And the questions "Why is he trying to escalate the war?" or "Why is he acting on that belief?" are different from the question "Why is he writing the letter?"

Perhaps we feel that the act was not done voluntarily. But purposive acts can be involuntary. For example, a man can be coerced to reach into his wallet for money. So the feeling we have about the two cases may have to do with matters other than purposiveness.

I shall nonetheless try to emasculate this objection by showing that any feature of these cases that leads us to suppose that the acts are not purposive can be referred to in a version of the causal thesis.

Evidently a person can do *A* in order to effect *B* without being able to specify all the conditions of his doing *A*. He may, for example, be ignorant of certain neurological requirements. So the fact that Jones cannot specify a sufficient condition of the act of writing the letter is not the feature that inclines us to say that he is not acting purposively.

Nor is the fact that another human being has produced the conditions that, together with existing conditions, add up to a sufficient condition of the action. For this feature is present when one human being commands, convinces, persuades, or brainwashes another human being to do something. And surely when Smith convinces Jones he should write the letter, Jones writes the letter in order to escalate the war.

Perhaps what disturbs us then is the way in which the hypnotist has bypassed Jones's psyche. Smith had to take Jones's intelligence, personality, and belief system into account in order to convince

¹⁴ In conversation.

him to write the letter; the hypnotist did not. This feature would also explain why we would be equally disturbed if the effect on Jones had been produced by physical manipulation of his brain rather than hypnosis.

The hypnotist (1) may have induced the pertinent desire in Jones without regard to Jones's present intelligence and personality (except certain minimal conditions like Jones's ability to understand the hypnotist's directions) or (2) he may have caused Jones to write the letter without inducing the desire to escalate the war in him. The reason other techniques of personality change, e.g., psychoanalysis, do not disturb us in the same way is that they must work on present personality and intelligence in order to bring such results about.

(1) is vague. Suppose I tell Brown that he should go to 412 W. 113th Street because they blop there and it's really great. Now Brown does not know what blopping is; but he takes my word for it (he need not even generally trust me) and forms the desire to blop or to do whatever I was referring to. This desire was induced independently of his personality and intelligence. He cannot perhaps be highly cautious or distrustful; but these kinds of character traits might be necessary for the effectiveness of hypnosis or brain surgery. Surely, though, when Brown goes to 412 W. 113th Street, he is going to blop.

Perhaps then (1) should be replaced by (3): The action is directed by a desire that was induced non-rationally, i.e., by means other than persuasion or the advancing of arguments, etc., whether the reasons and arguments be good or bad.

Besides the fact that the distinction between rational and non-rational inducement is very vague once we leave the arena of clear cases, there is the obvious objection that innately-based desires or needs are not produced rationally although a man seeking food, say, is surely acting purposively.

(3) may have to give way to (4), therefore. Perhaps there must be an essential reference to another person in the genesis of the desire. We are not bothered about innate desires. But we might be bothered about the creation of desires by another person, e.g., the hypnotist. In other words, we are disturbed about manipulation, not causation *per se*. If a man's brain happened to be in the state that the hypnotist produced in Jones, we should not be disturbed about that man's purposiveness if (4) is the answer.

In order to define (4), we must again distinguish manipulation from persuasion, argumentation, and even coercion, for these latter do not preclude purposiveness. Perhaps (3) provides us with a clue. Perhaps the creation of a desire for x in Jones by a hypnotist or brain surgeon is manipulation because it does not depend on the presence of formation of a belief in Jones that x would be satisfying in some way. This type of belief was necessary in order for Brown to form the desire to blop, but is not necessary when the desire is not created by another person, e.g., an innately-based desire.

Suppose, however, that the hypnotist creates both the desire to x and the belief that x would be satisfying in some way. Again we may feel that the hypnotist is the only "real agent" in the situation, especially in Case II where he also creates in Jones the belief that his action will have the desired effect.

When a man creates in another certain desires or beliefs, but is not engaged in manipulation, there is the assumption that he is constrained to some extent by the man's present intelligence, capacities, personality, and belief system. Although we have failed to define these constraints, it seems to me that this conclusion constitutes an emasculation of the objection. For the kind of condition we would have to add to the definition of purposiveness would say roughly that if P 's action has a desire or belief as a condition, and if either is created by another person (or person-like entity or entity believed to be a person if Jones can be induced to perform a non-purposive action by a robot whom Jones believes is human), then at the time of creation, other conditions of the creation are aspects A , B , and C of P 's intelligence, capacities, personality, and belief system. I see no inroads here for primitivism.

Restrictions on the power of an agent other than P would also be required to reply to an objection similar to the present one.¹⁵ Suppose God or some powerful being simply makes it the case that P does A whenever it happens to be the case that P has the appropriate belief—and, perhaps, other conditions are satisfied. Again, it is difficult to say exactly how we should formulate the appropriate restrictions. But it seems clear that these restrictions will not imply or suggest primitivism.

We have not yet dealt with (2), the suggestion that the disturbing feature is the action's not being directed by any desire or reason, although caused by the hypnotist (a situation that may also hold in the God case). If purposive actions have to be

¹⁵ M. Slote mentioned this problem in conversation.

V. ESSENTIALISM AND THE SENSES OF PROPER NAMES

GERALD VISION

I

IT has been argued, contrary to widely accepted doctrine, that

- (1) Proper names have a sense formed by a sortal property¹ in a way that other expressions in a language, particularly common nouns, can be said to have a sense.

This has been maintained in close conjunction with a corresponding doctrine about individuals that states, contrary to the orthodox interpretation of Locke, that

- (2) Individuals have (nominal) essences.

The discussion of (1) and (2) below is restricted, unless otherwise specified, to those forms of either doctrine that can be construed as providing support for the other one. In those forms it is either made to appear that one of the doctrines rests on the other or it is held that they are alternative ways of stating a single doctrine. For the sake of brevity I shall call (1) and/or (2) essentialism, realizing that this is not customarily the title of a doctrine concerning the senses of words and that I have not exhausted everything philosophers have understood by this term.

Given the connection with (2), the sense to be discussed in relation to (1) is not that associated with any morphemic element we might find in proper names, but is, quite specifically, a sense that would have to be provided by the kind of the individual designated by the name. A brief canvass of the kinds of individuals which bear proper names would include not only men and domestic animals, but also buildings, constellations, battles, and landmarks. The vast differences between these may make one justly suspicious of a generalization such as (1) or (2), and—if the intelligibility of the

doctrines be allowed—equally suspicious of contrary generalizations. Therefore, I shall concentrate on essentialist claims concerning spatially contiguous particulars which characteristically persist more than a brief moment: especially human and other animals. It seems safe to ignore problems that have their source in the differences between types of individuals since the essentialist arguments I wish to discuss are designed with the aforementioned type of case in mind.

II

From the close connection between (1) and (2), it might be argued that if a proper name has a sortal concept connected with its application, the exemplification of which is a necessary condition for applying that name to its bearer, this must be reflected in a corresponding property which that individual possesses essentially. Or similarly, if an individual possesses a sortal property essentially it would seem that this must be reflected in any proper name of that individual by having that property determine (part of) the sense of the proper name. But I do not believe that correspondence between a single proper name and a single individual *qua* individual or the parallelism of the relations between a proper name and its sense and an individual and its nominal essence can be so easily assumed.² It seems open to further argument that (1) or (2), but not both, may be true. One conceivable—though not necessarily more than merely conceivable—state of affairs that would permit (1) to be true while (2) is false is a case where proper names are used to designate individuals only *under certain restrictions*, and not *per se*. This suggests that a proper name is no longer applicable to an individual after certain alterations which the individual

¹ I use "sortal property," "sortal concept," or simply "sortal" throughout this paper roughly following Strawson's use of "sortal universal" in *Individuals* (London, 1959), pp. 168 ff.

² As in John R. Searle, "Proper Names" in *Philosophy and Ordinary Language*, ed. by Charles E. Caton (Urbana, 1963), where it is stated without formal argument that "What is Aristotle?" and "What are the criteria for applying the name 'Aristotle'?" ask the same question, the former in the material mode, and the latter in the formal mode of speech." (P. 159.)

could nonetheless survive-identity intact.³ If this is not the way in which proper names are in fact used to designate individuals, this must be shown and not assumed.

There are replies to these considerations in the offing; but tracing them would take us too far astray from our main topic. So I shall assume we are discussing a restrictive notion of *senses* bringing them in line with nominal essences. Briefly, let us suppose that ' Φ ' is a sortal term expressing the property ϕ , and ' X ' is a proper name designating an individual x , such that ' Φ ' cannot be a part of the sense of ' X ' if x could cease to be ϕ (even momentarily). Let us now examine whether this restricted type of sense is possible for proper names—or, alternatively, whether its corresponding nominal essence is possible for individuals.

III

The remainder of this paper will be devoted to the examination of two lines of argument that essentialists have recently used to support their contentions. The first of these arguments, hereafter (I), is stated in a very compressed manner by G. E. M. Anscombe. She argues that

The doctrine that individuals have nothing that is essential to them suggests a phantasmic notion of the individual as a "bare particular" with no properties, because it supposes a continued identity independent of what is true of the object.⁴

If an individual has no nominal essence, there is no property the loss of which it cannot survive. And, if proper names have no sense, the individuals they designate could not have nominal essences. But if an individual can survive the loss of any property whatsoever, it must be something independent of any property whatsoever. Thus, it must be a bare particular. Since the notion of a bare particular is logically unsound, all individuals must have nominal essences, and expressions of these essences will naturally provide the senses of the names of the individuals.

It is unnecessary to decide here whether the conception of a bare particular is, as alleged, phantasmic. For it can be shown that the denials in

question do not involve bare particulars understood in a way that has drawn all the objections. The doctrine of bare particulars can be represented adequately for our purposes as affirming,

- (3) $\Diamond(\exists x)[(\exists y)(x = y) \ \& \ \sim(\exists \phi)(\phi x)]$.

An identifiable individual having no properties is conceivable or possible. (Or, it is possible that there be an individual identifiable independently of any of its properties.)

(I omit presently irrelevant refinements, such as the need to add restrictions to the substituent set for ' ϕ ' in order to avoid a logically false claim.) However, one way to represent the claim that an identifiable individual has no properties that are essential to it is,

- (4) $(\forall x)[(\exists y)(x = y) \supset \sim(\exists \phi)\Box(\phi x)]$.

There is no single property that is necessary for the identity of an (arbitrarily chosen) individual.

(4) does not entail—or "suggest"—(3): therefore the consequences Anscombe foresees do not follow.

We could have denied essentialism by asserting

- (5) It is not necessary for any or some identifiable individual that it have some property or other (where "some property or other" amounts to "any property").

Any adequate construal of (5) might be incompatible with the denial of (3). But if (3) is an obviously objectionable doctrine, and (5) entails (3), this would seem sufficient reason to offer the opponent of essentialism something on the order of (4) rather than (5). (4) is compatible with an individual changing all of its properties and retaining its identity; or a more austere version still compatible with (4) would be that an individual could retain its identity through a change of *any*, though not necessarily *all*, of its properties. Arguments examined later may exclude (4). The point of the present argument is that (4) is not excluded by (I).

Another way to highlight the inadequacies in the essentialism of (I) is to trace the consequences of rejecting (3); for this is as strong a commitment as

³ A version of an essentialist argument offered by P. T. Geach, "Good and Evil," *Analysis*, vol. 17 (1956), pp. 33-42, overlooks the possibility of a lack of correspondence between the relations exemplified in (1) and (2). There he says: "... the continued use of a proper name ' A ' always presupposes a continued reference to an individual as being the same X , where X is some common noun; and the ' X ' expresses the nominal essence of the individual called ' A '." (P. 34.)

⁴ G. E. M. Anscombe, "Substance," *Proceedings of the Aristotelian Society*, Supplementary Volume XXXVIII (1964), p. 70, my italics. There is the barest hint of support for this line of reasoning in David Wiggins, *Identity and Spatio-Temporal Continuity* (Oxford, 1967) when he asserts "My essentialism simply derives from a willingness to pay more than lip service to the idea that we cannot single out bare space-occupying matter." (P. 43.)

(I) requires. Obviously, it will not be enough to assert that an individual must always have some properties or other, which is all the rejection of (3) enjoins. But let us even suppose that it is required that an identifiable individual necessarily have one property or set of properties. It may be thought that this supplies the deficiency. But if that property cannot be further specified, and there is no hint in (I) of any device for specifying it, it is of no moment for us. This may seem obvious in the case of (1), for the sense of a word which cannot be specified by a user of that word is not the sort of sense that could help the user to discern correct applications of the word; and if a sense does not provide direction for the application of an expression, how could it really be a sense? If we allow that a nominal essence is a type of convention, it seems the inability to specify that essence is a strong presumption against the intelligibility of (2) as well. Certainly, we may say at a minimum that even if this is an intelligible doctrine of nominal essence (without any further premisses), it is not the kind that will have a bearing on the senses of proper names.

IV

A more popular essentialist defense places central reliance on what I shall call *the sameness principle*. Though the principle is more frequently employed as a part of the argument below numbered (IIa), another argument using a similar version of this principle, (IIb), must be considered in conjunction with it. For there is some evidence that applications of the sameness principle in (IIb) are smuggled in to remove deficiencies in (IIa): though Geach, whose *Reference and Generality* is my source for the argument, only hints at (IIb), and never adopts it explicitly.

The sameness principle is the claim that phrases of the form "the same . . ." are logically incomplete and demand completion by a count noun or substantival expression of relatively restricted applicability. (In the material mode: two things are

never simply *the same*, but are always the same *something-or-other*.) I argue below that the conjunction of (IIa) and (IIb) to yield the essentialism embodied in (1) and (2) is illicit, since the substituent set for 'X' in "the same X" should be differently determined for each argument. After arguing this, I attempt to show that (IIa) alone will not force us into essentialism. Now let us summarize both arguments.

(IIa) *The identification argument*. If one knows how to use a proper name, sense must be given to the question whether the individual designated by the name on one occasion is or is not identical with an individual designated by the name on some other occasion. We could not say that one knew how to use the name unless one could answer this question correctly in a large proportion of certain types of situation. This amounts to knowing when the individual is the same or not. By the sameness principle we require that the individual must be the same (or a different) X, where the substituent set for 'X' includes a subset of common nouns (Geach), substantivals (Geach), terms expressing sortal concepts (Wiggins), or, alternatively, count nouns (Geach; Cartwright). The concept under which the individual is identified provides the sense of any proper name designating that individual; for any necessary restriction in the identification of an individual is naturally reflected in the rules for the application of its name. Furthermore, if having such a property is a necessary condition for identifying an individual through time, it seems that no clear sense can be attached to the suggestion that the individual could retain its identity after the loss (or before the acquisition) of that property.⁵

(IIb) *The substitution argument*. If there is an expression ϕ and there is another expression or series of connected expressions ψ , and substituting one of the series of ψ for every occurrence of ϕ allows the speaker to convey conventionally all the information conventionally conveyed using ϕ , then the substantive part of ψ expresses the (whole) sense of ϕ .⁶ At any point in the identification of an indi-

⁵ I take this as a summary of the type of argument given by G. E. M. Anscombe, "Aristotle" in G. E. M. Anscombe and P. T. Geach (eds.) *Three Philosophers* (Oxford, 1961), pp. 10-11; P. T. Geach, *Reference and Generality* (Ithaca, 1962), pp. 43-45, and *Mental Acts* (London, 1967), p. 69; and outlined—though not defended—in Richard Cartwright, "Some Remarks on Essentialism," *Journal of Philosophy*, vol. 65 (1968), p. 624.

⁶ Though Geach suggests this line of argument on p. 45 of *Reference and Generality*, he also warns against *the cancelling-out fallacy* which amounts to the position that if two propositions verbally differ precisely in that one contains the expression E_1 and the other the expression E_2 , then, if the total force of the two propositions is the same, we may cancel out the identical parts and say that E_1 here means the same as E_2 . (P. 61; *ibid.*) It does not seem that the fallacy is committed in this argument because the principle to which this footnote is attached is quite general—requiring replacement in all contexts, and not just on some occasions. It is a principle, in other words, concerning "means the same as" and not simply "means the same as *here*." That Geach did not intend the fallacy to apply to such general principles is at least indicated by his choice of examples following his enunciation of it.

vidual we must be able to say *what* the individual has been for the entire span of its existence up to that point. Thus, if it is Socrates who is being identified, we may say that the individual is *a man* or is *the same man* as such-and-such. Since proper names name individuals of certain types, it would seem that they are (theoretically) replaceable by the original identification of an individual as of a certain type—say, a man—and repeated uses of phrases such as “the man” or “the same man” anchored by the original identification. Therefore, proper names are instances of ϕ for which count nouns plus “same” plus applicatives (e.g., “a,” “the,” “any,” “no,” “every,” “some,” “just one,” “many”) are ψ . (A more informal way of conceiving the argument would be to raise the question “If there were no proper names in the language, what linguistic element if any could we use to convey the information we now convey linguistically with proper names?” The answer to the question, if there is one, provides us with the material for senses of our proper names.)

V

Let us first examine the sameness principle in (IIb). Though it may turn out that any count noun that satisfies an application of the sameness principle in (IIb) can be discovered to satisfy an application of it in (IIa), we are nevertheless presented with a wider choice of count nouns for any application of the principle in (IIb). This is because (IIa) requires that the count noun be used, and usable, on the given occasion to identify the individual in question; while (IIb) only requires that the property expressed by the noun be true of the individual at some arbitrarily selected stage within the individual's existence. If every such stage in the existence of an individual meets this requirement, the sameness principle, and hence the argument, is satisfied. For example, at T_2 (where T_2 is later than the time of the individual's inception, T_1 , and earlier than its termination, T_n) there must be some count noun which will characterize the individual in question from T_1 to T_2 . Similarly, there must be some count noun characterizing the individual from T_1 to T_3 . But there is no logical presumption in favor of the two count nouns expressing only one concept between them. If the count noun at T_2 is “boy,” while at T_3 or T_n it is “person,” this may seem

obvious. But I shall argue for the stronger claim that this is possible where the first kind is neither a species nor a normal stage in the development of the second kind. Also, from the requirement of (IIb) alone, the use of a certain count noun at T_{n-2} does not commit us to being able to apply that same count noun to the individual at some later time, T_{n-1} , as long as some other count noun is *available* to cover the individual from T_1 to T_{n-1} .

What makes a count noun “available” for use in (IIb)? Taking an example essentialists would clearly exclude, we might ask “Do we have available count nouns covering the various transformations ascribed to Proteus, who, it is said, turned from a man to a lion to a snake to a panther and, finally, to a bear?”⁷ (Let us leave aside for the present the claim that Proteus could become water or fire.) Nothing provided by (IIb) or its implications prohibits us from allowing that Proteus, through all these transformations, is the same *animal* or *creature*. If we lacked words in our vocabulary to cover all of Proteus' transformations, (IIb) alone would not prohibit us from coining new count nouns for the occasion—say, count nouns that definitionally amounted to “man or lion or snake or panther or bear.”

VI

At least two proponents of the sameness principle, Geach and Wiggins, offer some account of how to restrict and determine the legitimate replacements for ‘X’. Since these may affect the legitimacy of the innovative procedures of the last paragraph, we must now examine them briefly.

Geach limits the substituent set for ‘X’ to what he calls *substantivals*. A substantival is a type of general term for which the grammatical criteria are only a rough guide. His original distinction between *substantivals* and *adjectivals* seems based on the fact that substantivals “have (singular and plural) numbers on their own account”⁸ while adjectivals do not. This would make the difference reside in whether or not the concept has an arithmetic. Grammatically clear cases of substantives, such as “sea,” are ruled out because “. . . nobody could set out to determine how many seas there are; the term ‘sea’ does not determine any division of the water area in the world into seas in the way that the term ‘letter’ . . . does determine a division of the

⁷ Homer, *Odyssey*, Bk IV: 453–464.

⁸ *Reference and Generality*, *op. cit.*, p. 39.

printed matter in the world into letters.”⁹ I said only that this *seems* to be Geach’s distinction because he shortly admits mass nouns such as “gold” into the category of substantivals on the ground that we can talk of the same gold. But I shall ignore this exception, as well as its revised criterion for substantivalhood more closely tied to “identity” than “countability,” for the following reasons: (a) it is inconsistent with the exclusion of expressions such as “sea”; (b) it would require that we accept considerations based on the necessary conditions for identifying individuals, and this would have to be an argument something on the order of (IIa); (This has a twofold drawback. On the one hand, the sameness principle is a premiss of (IIa), and therefore cannot be proved by (IIa). On the other, our investigation of (IIb) cannot admit the inclusion of considerations from (IIa), and thus we are told nothing about the sameness principle as it applies to (IIb) alone.) (c) it appears that Geach later gives up the revised criterion for one in terms of count nouns—for he claims in a later work that all we need admit for identity are “. . . as many as we need of two-place predicables of the form ‘— is the same *A* as —’ where ‘*A*’ is some count noun.”¹⁰

This seems to be the extent of Geach’s account of present concern, but a closely associated doctrine appears in David Wiggins’ *Identity and Spatio-Temporal Continuity*. He seems to agree that ‘*X*’ is replaceable by any term expressing a sortal property, but there are two kinds of sortals: substance-sortals (Wiggins calls them *substance-concepts*) and phase-sortals. (I shall refer to their linguistic correlates as substance-terms and phase-terms.) As initially distinguished substance-sortals are those “. . . which present-tensedly apply to an individual *x* at every moment throughout *x*’s existence.”¹¹ He gives as an example *human being*, and it appears he would readily agree with Geach’s additional examples *cat* and *dog*. By implication, phase-sortals are those which do not apply to *x* at every moment of *x*’s existence. Examples are *boy* and *cabinet minister*. Substance-terms are clearly allowable substitutions for ‘*X*’. Wiggins is never as clear as one could hope about phase-terms substituting for

‘*X*’: but what is certain is that if a phase-term is a substitution for ‘*X*,’ there is always a substance-sortal which the phase-sortal it expresses *restricts*. For an account of what it is for one sortal to restrict another Wiggins refers us to Geach’s *Reference and Generality*. For Geach, a substantival ‘*A*’ is a restriction of another substantival *A* if ‘*A*’ can be put in the form (i.e., amounts to) “*A* that is *P*” (where “is *P*” is a predictable). It would seem that any sortal which could be glossed *per genus et differentia*, including substance-sortals such as *man* (human being) or *cat*, would be restrictions of some other sortals. Therefore, it is puzzling that Wiggins, in his next elaboration of the distinction, without any further argument on the subject, remarks that it is a distinction between substance-sortals on the one hand, and “restricted or phase-sortals” on the other.¹² It is likely that he has in mind a narrower conception of *restriction* whereby a restricted sortal represents one stage in the development of a more broadly conceived individual. (The stage need not be a normal one: e.g., *cabinet minister* is a restriction of *human being*, though not a stage that most humans go through.) But if this is taken to advance our knowledge of the distinction it is misleading. For to explain a *restriction* as Wiggins understands it we must employ the concept of a sortal that applies to an individual at some but not every moment of that individual’s existence. And this is how the distinction was initially drawn.

With the failure of *restriction* to provide further clarification there are still puzzles about how the distinction is to be construed. Is *boy* a substance-sortal for an individual who dies in childhood? What about *king* for an hereditary monarch who is born to and dies on the throne? Also, how do we separate these sortals? Must we, for example, wait until the individual ceases to exist to know whether *human being* or *man* is a substance-sortal? Thus far, nothing in our directions for using this distinction prohibits our makeshift procedure employed in explaining how Proteus could change so radically in connection with (IIb).¹³ Wiggins has offered *human being* as an example of a substance-sortal, and this would exclude Proteus’ transformations. But nothing in his account up to now justifies his

⁹ *Ibid.*, p. 38.

¹⁰ P. T. Geach, “Identity,” *The Review of Metaphysics*, vol. 21 (1967), p. 10.

¹¹ *Identity and Spatio-Temporal Continuity*, *op. cit.*, p. 7.

¹² *Ibid.*, p. 30.

¹³ Wiggins offers a dark hint that he would be willing to allow sortals coined *ad hoc* to meet extraordinary situations when he admits that “. . . the structure may be thought of as including all the sortal-concepts whose possibility is implicit in the principles of classification embodied in those sortals actually named and used which belong to the structure. . . .” *ibid.*, pp. 32–33. (My italics).

example. However, apparently just in elaboration of the original distinction, Wiggins introduces a new and stronger distinction based on whether " x is no longer f " entails " x is no longer."¹⁴ If the entailment holds, f is a substance-sortal: if it does not, f is a phase-sortal. So it seems as if being a phase-sortal or a substance-sortal is a matter determinable by the senses of the terms expressing sortals.¹⁵ But this assaults the nostrils piscanally, for it is obviously not the distinction Wiggins originally proposed. If the claim is now going to be made that *human being* is a substance-sortal more argument will be needed. Since Wiggins goes on more-or-less to assume that individuals do take substance-sortals, he has not simply introduced a distinction, but made a rather substantial philosophical claim, practically amounting to affirming essentialist thesis (2). If we previously doubted that individuals have nominal essences it will now take the form of doubting that individuals exhibit substance-sortals (viz., that the sortals they do exhibit are substance-sortals).

VII

In discussing (IIa) it seems best to concede a central role for some sortals in determining persistence conditions for individuals which instantiate those sortals. Sortals are also connected with our other resources for identification in other systematic ways to be explained below, and I shall take notice of this by the following concession, called *the weak thesis*:

- (6) For a person, P , to have a criterion of identity for an individual, a , P must have a sortal concept, X , under which P may be confident that he can identify and re-identify a .

The weak thesis does not entail:

- (7) " $(P \text{ identifies } a) \supset \Box (Xa)$ " or " $\Box [(P \text{ identifies } a) \supset (Xa)]$ " or " a is necessarily- X " or "It is necessary that a is X ."

One can be confident of p , be justified in being confident of p , and it may still be the case that not- p . I shall call all interpretations of the sameness principle that require some truth along the lines of (7) *the strong thesis*. Since a holder of the strong thesis is likely to accept (6) as well, it should be made explicit that hereafter, unless otherwise

specified, when I speak of someone holding the weak thesis I shall mean holding *only* the weak thesis (viz., assenting to (6) and rejecting (7)). The weak thesis allows X to be a criterion for identifying a in the idiomatic sense of "criterion" (i.e., something we can go by), but allows a to persist even where it is not X . But why should we prefer (6) without (7)?

VIII

A wealth of common instances of biological transformations appears inconsistent with the strong thesis. Individual caterpillars (Larvae) become either butterflies or moths, individual tadpoles turn into frogs or toads. It seems more natural to trace a single individual through successive stages than to say that the infant individual has ceased to exist and another individual—what would otherwise be the mature stage of the original individual—has come into existence. (Think of the case where a father is explaining to a child about the sudden absence of his pet tadpole and the presence of a frog.) The transformations are radical. "Caterpillar" and "tadpole" are count nouns under which individuals could be identified. And, for the case of caterpillars at least, the transformation takes place in a hidden medium so that it might not seem gradual to a relatively uncasual observer. The reply that all this shows is that "caterpillar" and "butterfly" do not belong to the substituent set for ' X ', just as "boy" does not, and that we need another count noun, say "*Lepidopteron*"; under which we may continuously identify the individuals in question, only pushes the problem back a short step. For *only if* it can be independently decided that individuals of type X are identical with individuals of type T can a count noun "restricting" (in Wiggins later sense) both X and T be legitimate. And this is just the situation the essentialist—holding the strong thesis—claims can never occur if we have identified individuals as X or as T .

The point of the case emerges more clearly when we have previously identified individuals—as the case of the elvers called *Letpocephali*—that are thought to be distinct creatures, but later found to be a stage in the development of differently identified individuals—in this case *Conger Eels*.¹⁶ How is this discovery possible? Before the discovery each count noun had the status we roughly ascribe to

¹⁴ *Ibid.*, p. 30.

¹⁵ Though Wiggins appears to issue a disclaimer at several places, including *op. cit.*, p. 69, note 37.

¹⁶ *Ibid.*, p. 59.

“man” (or “human being”) and “cat,” and the strong thesis seems to rule out our ever discovering that a man has become a cat. Certainly, the essentialist could not consistently admit this as *fait accompli*.

IX

David Wiggins has attempted to describe the examples of metamorphic change (offered in VIII) in terms presumably compatible with essentialism.¹⁷ He claims that aside from our two classes of sortals, substance- and phase-sortals, there are sortals that are “porous” or “indeterminate.” By this he means that “they enable us to pick out (things falling under them) during some stretch of their existence,” but “leave quite open the character of (those things) during other periods of their life history.” They are distinguished from phase-sortals because phase-sortals are restrictions of other sortals, ultimately of substance-sortals. All substance-sortals are *determinate* in that they lay down, sometimes in a general and sketchy way, the boundaries past which an individual of that type can no longer exist. With the advance of knowledge we may make porous sortals restrictions (viz., phase-sortals) of more determinate sortals. We are allowed to continue to identify an individual only falling under a porous sortal through the change of that property partly because of the indeterminate character of porous sortals described above and partly because we have “the *general* concept of a continuant through change,” which need not be determined by a particular sortal concept. One can then

... if he wishes (or, if he is constrained by analogy with what he already does in other cases) ... count a suitable event or process which he witnesses befall an *f* as the discovery that *fs become f's*.¹⁸

The account should increase our suspicion that all the examples Wiggins has given of substance-sortals are really only porous sortals. A porous sortal, say *Leptocephali*, as introduced before we discovered that *Leptocephali* were only stages of Conger Eels, is certainly a category under which

individuals were identified and under which we could trace the coincidence of two individuals. These are just the roles that Wiggins has previously argued made a concept a substance-sortal.¹⁹ If adopting these roles does not prohibit the sorts of changes that would be incompatible with essentialism, Wiggins has forfeited his reason for saying of his examples of substance-sortals that they cannot allow these types of changes to take place. It looks as if in all relevant respects the teaching of a sortal-term such as “human being” is indistinguishable from the teaching of an expression for a porous-sortal. In both cases one is likely to choose clear-cut instantiations of the property to give as examples and to list characteristic features of an individual of that type. How, then, are we to decide when the persistence conditions are determined and when they are not?

On Wiggins’ account the “discovery” that *Leptocephali* are infant Conger Eels is not really an empirical discovery in the most favored way. It is partly infected by a *decision* to change (by determination) the former sortal concept, and can be carried out only *by analogy* with what goes on in full-blooded empirical discovery.

Though there is nothing preventing a redescription in these terms—indeed, it might be carried out for *any* case of empirical discovery—it obscures what we want to stress: namely that the change is a result of greater empirical knowledge. To represent us as juggling our conventions (viz., the senses of our sortal-terms) in any more prominent way than happens in the case of any empirical discovery hides just that feature of the case which seems most important.²⁰ Unless Wiggins can bring strong grounds independent of the thesis he wishes to salvage in support of this redescription of the situation, I can see no reason why we should favor this out of all cases of putative empirical discovery for redescription.

X

We have yet to see why count nouns deserve the central role in our identifying procedures conceded them by (6). Furthermore, there may be lingering

¹⁷ *Ibid.*, pp. 59–60; p. 69 (note 37).

¹⁸ *Ibid.*, p. 59.

¹⁹ *Ibid.*, pp. 34–36.

²⁰ This rebuttal is a close paraphrase of a point made by Hilary Putnam in “Dreaming and Depth Grammar,” *Analytical Philosophy*, ed. by R. J. Butler (Oxford, 1965) p. 220. Putnam was talking about what may at first appear to be a completely different topic; but there may be more analogy between the expressions he was discussing and proper names than we at first imagine. (Cf. pp. 218–221 of Putnam’s article.)

discomfort if it is thought that my examples undercut any clear distinction between *a*'s going out of existence (being replaced) and *a*'s persisting through change. The essentialist avoided blurring that distinction by insisting that if *a* and *b* are the same, they must be the same *X*. I shall attempt to justify (6) and abate the discomfort by making a few desultory remarks about identification procedures in general. These in no way constitute a theory or systematic account of identification. I shall be content if I have explained how and why it is possible to abrogate the use of a sortal when confronted with extraordinary situations, where that same sortal served as the basis for identifying a presently considered individual.

If the sameness principle represents a requirement no stronger than what is necessary for its application in (IIb)—viz., that we could always find or coin some sortal to cover all previous stages of an individual—we can accept it. Furthermore, we can allow something to the claims of the sameness principle as used in the identification argument, (IIa). We can admit that since individuals do not regularly change the sortals under which they are identified, the use of such sortals in garden variety cases of identification is warranted and advisable. (The counterexamples were only drawn to show that what is ordinarily true or always useful is not thereby guaranteed *a priori*.) Also, if sortals were our only means for reidentifying an individual, we might have been prevented from claiming that they could ever be abrogated in identifying situations. But it is quite obvious that a sortal concept, say *man*, is not the only thing involved in identifying and reidentifying individuals of a certain type: for to tell that *a* is identical with *b* we must know more than that *a* is a man and *b* is a man. We must know that *a* and *b* are *the same man*: in Wiggins' terms, that they *coincide* under the concept *man*. And sameness or coincidence is not wholly imparted by the sortal concept covering it, else knowing that both were men would be enough to identify them. A further consequence of this unsatisfactory doctrine would be that no change which befell *a* could terminate *a*'s existence if the result of the change was still a man.²¹

From these considerations it is obvious that there must be tests for identity other than the sortal concept under which an individual falls. In the case of

human beings no doubt memory and character traits play a supplementary role that they cannot play with other concrete individuals. But for large groups of concrete individuals spatio-temporal continuity is a very important, though perhaps never sufficient test for identity. A frequently overlooked factor in deciding whether an individual has changed or has been replaced is the (relative) gradualness of the change. This will never be more than a supplement to other tests of greater importance; but it is conceivable that we might make different judgments on two occasions where the only relevant difference was the gradualness with which the transformation took place. (E.g., suppose the caterpillar went up in a puff of smoke and a butterfly appeared at its former place.)

Most of the time the other tests listed will yield results consistent with the application of the sameness principle. This is not fortuitous, for to some extent what will count as, say, spatio-temporal continuity is determined by the *kind* of thing in question. But it cannot be concluded from this that it is not possible to apply the test for spatio-temporal continuity independently of the sortal under which the individual is originally identified. Taking the example of animals, each species does not lay down conditions for the continuity (and persistence) of its members that are peculiar to that species and very different from those of all other species. Therefore, *a priori* claims to the effect that the tests for identity must yield consistent results are not supported by this type of consideration. It seems possible in the relatively rare cases where, on analogy with the common cases, a creature is spatio-temporally continuous with what was in a certain place before, but we cannot say of both creatures that they are the same *X* (because '*X*' is not applicable to the more recent creature), we may identify the distant and recent creature as the same *X* instead of saying that one individual terminated and a new one occupies its former space. And this might be an accurate description as far as it goes of what happened in the examples of metamorphoses cited.

This in no way impairs the claim that the identified individual must always be subsumable under some sortal or other (the conclusion of IIb.). For identifying involves reidentifying and, at times, individuating, and these operations often require

²¹ Wiggins would not accept this consequence. The example from Hobbes he discusses (*op. cit.*, p. 37), where a plank-hoarder reconstructs a ship from discarded parts of a particular ship, would make the reconstructed parts a ship, though, he holds, not one identical with the original ship. He agrees that "ship" is a substance-term.

counting.²² And counting is only possible for *kinds* of things.

It may be apposite to see how this bears on Proteus' alleged transformations into water and fire. Such transformations may indeed be unintelligible, but if so it would appear that the reasons would have to be more complicated than those given in (IIa). Aside from the sortal we employ much depends on the details of our description in individual cases. We must know not only if the change was gradual or the fire could emit noises resembling sentences that we could have expected Proteus to utter; but, also, whether Proteus could announce a putative intention to perform the transformations, and whether this could be regularly repeated or only happened once. We must also consider whether, once the cycle of changes is complete, we have before us again something resembling the original man Proteus, or whether the fire cannot be replaced by or retransformed into a man. If, at the end of the cycle we have overwhelming evidence that the man who appears is indeed still Proteus, we will not have avoided altering our identification procedures. For even if we refuse on *a priori* grounds to say that the fire is identical with Proteus, we shall be forced to allow that we now have individual men who can come into and go out of existence at intervals, and this too is a type of change in our original sortal *man*.

It is of no avail to suppose that the animal-like transformations of Proteus are intelligible and the ones discussed in the preceding paragraph unintelligible because we have the count noun "animal" to cover the former, and no genuine count noun to cover the latter. For if it is the sortal that is supposed to aid in identification by preparing us for normal changes in the individual, we must take account of the fact that *there is no such animal* as a man at one time, a lion at another, and a bear at yet a third. Therefore, using the sortal *animal* could not be of any help in identifying Proteus through all these transformations. In other words, the problem is not essentially one of having a term to characterize the case; for even if we have such a term, and previous use justified us in adapting it to cover our present case, the new employment forces on us the same sort of innovation which would have been required had we to coin a new term appropriate to the unusual circumstances.

What can be said of the role of sortals in identifying individuals is not inconsiderable, though, as I hope to have shown, it is not enough for the essentialist. Though the essentialist's general recipe for distinguishing termination (or replacement) from mere change relied on the position just criticized, it should not be concluded that the distinction is now either obliterated or arbitrary. None of the other single tests mentioned in this section are of comparable decisiveness. But even if we find no others, collectively they provide us with a number of objective considerations on which to base our identifications and reidentifications.

XI

What the essentialist had hoped to elicit from (IIa) was a rather strong doctrine demanding relatively specific sortals such as *man*, *cat*, *building*, etc., for the nominal essences of individuals. But how strong a doctrine one is entitled to depends on the specificity of the substituend set for 'X' in the sameness principle vital to applications of (IIa). Therefore, the choice of a substituend set should not itself depend on the outcome of argument (IIa). This brings out another defect in the line of reasoning being discussed which, I believe, is made clearest by the following hypothetical exchange. What could the essentialist reply if I chose "animal" rather than one of its species as the substituend for 'X' in the sameness principle? He might argue that the need for a more specific substituend is dictated by what is necessary for the identification, etc., of an individual.²³ But how could this answer be correct, for surely if I say truly of some individual *a* that it is the same animal as *b*, I have correctly identified the individual. The correctness of my identification is determined by the correct application of *the same*, no matter how general or specific a legitimate sortal I attach to it. And *animal* is a legitimate sortal, since animals can be counted, which is a sufficient condition for being a substantial.²⁴ The obvious reply for the essentialist at this point is that, say *man* is a much more useful concept to use for reidentifying something than *animal*, for it yields more information. This is correct, but it leads to the question why could we not choose something specific enough to be of optimal use, such as "the same boy" or "the same locomotive engineer"? Has the essentialist

²² But see Geach's counterexample discussed in note 6.

²³ See Wiggins requirement (D. vii), *op. cit.*, p. 37.

²⁴ Geach, *Reference and Generality*, *op. cit.*, p. 39.

any answer except that one individual can change from being a boy or a locomotive engineer, but cannot change from a man to something else? But what proof have we of this aside from the argument which requires this principle for soundness?

XII

It may be conceded that (IIa) demonstrates that part of the apparatus anyone who is said to have a criterion of identity for an individual must have is a sortal concept under which he (the identifier) can be reasonably confident of identifying the individual. But this is no assurance that the concept will be applicable on every occasion that he successfully identifies the individual. This is the weaker thesis. But it may be combined with the conclusion from (IIb), which requires that there is always some sortal under which the individual is successfully "identified," to yield the spurious doctrine that the sortal one has in his criterion for identity of *a* must be successfully applied if *a* is to be identified (or have an identity). This amounts to the strong thesis. I call it spurious here, for the sortal which is demanded by the conclusion of (IIb) need not be identical with the one mentioned in the premisses of (IIa). To see this we need only remember the results of our discussion of (IIb), where it was argued that the sortal demanded there for the successful application of the sameness principle may not enable us to identify the individual in advance, but, quite the contrary, may on occasion be the *result*—rather than the *cause*—of our ability to identify an individual.

XIII

The essentialism I have been discussing should be distinguished from another doctrine that could go under the same title. (Though a different doctrine, this other sort of essentialism has been used as a defense of (1) and/or (2) as illustrated below.) This second doctrine starts from the contention that though individuals can change sortals, the changes they can undergo are not unlimited. For example, though Proteus may be able to survive many changes, he could not become the Equator or a prime number or an instance of the relation *to the left of*. (I am neither affirming nor denying that any of these items are, properly speaking, individuals.) A nominal essence on this doctrine would not be a simple sortal property, but a (possibly infinite)

disjunctive set of properties. As it stands it seems that this doctrine of nominal essences cannot be used to support either a correlative doctrine of the senses of proper names or the rather specific contentions of some essentialists concerning the place of sortals in identification.

But might there not be some hope that we can coin a sortal covering all the permissible transformations of an individual in the way it was suggested such a sortal would cover the actual transformations of Proteus? I think not. To begin with, we must specify the nominal essence of an individual in terms of sortals that it could not take. Geach has claimed that the negation of substantial term is never a new substantial term,²⁵ and Wiggins has maintained both that a sortal tells us what a thing is, and that one "... cannot say what a thing is by saying what it is not."²⁶ If this is true of the simple negation of a sortal, could the situation be different with a disjunction of negations of sortals? If we coined a simple term to stand for the whole disjunction, it would not thereby be a sortal.

Ignoring the previous difficulty there are still problems in constructing a sortal that do not appear when the disjunction is known to be finite. For any finite disjunction of sortals it may be possible to construct a new sortal, and there is no problem of what to do in new cases not covered in a disjunctive definiens. But where the disjunction cannot be known to be finite, it must be supposed that the permitted (or prohibited) changes follow some pattern so that the sortal can generate a rule for an unbounded class of cases. But is there any reason for holding that the permitted or prohibited changes of individuals do follow some rule or pattern? And, if we did get a general rule which they proceeded in accordance with, would not some of the sortals turn out to be something on the order of "piece of matter," "space-occupier," "thing," etc.? If this is the consequence of our effort it is clear that our sortal will not support the doctrine concerning the senses of proper names the essentialist had desired to save. Nor would Geach or Wiggins accept these terms as legitimate substitutions for 'X' in the sameness principle.

In summary, the sort of essentialism introduced at the beginning of this section offers no prospect of finding properties at a level comparable to our ordinary species of animals to serve as essences.

Temple University

Received November 5, 1969

²⁵ *Reference and Generality*, *ibid.*, p. 40.

²⁶ *Identity and Spatio-Temporal Continuity*, *op. cit.*, p. 70, note 39.

VI. HUSSERL'S PROBLEMATIC CONCEPT OF THE LIFE-WORLD

DAVID CARR

AS Herbert Spiegelberg notes in his historical study of the phenomenological movement, "the most influential and suggestive idea that has come out of the study and edition of Husserl's unpublished manuscripts thus far is that of the *Lebenswelt* or world of lived experience."¹ Because this fertile idea has inspired so many original and insightful contributions to phenomenology since Husserl's death, notably in the work of Maurice Merleau-Ponty and Alfred Schutz, and since the investigation of the life-world seems firmly established as an important subject for philosophical concern, it may seem a matter of only historical interest to return to Husserl's writings for a critical analysis of his own thoughts on the subject. But philosophy, as Husserl recognized and insisted at the end of his life, is like any other cultural activity in existing as a cumulative *tradition*; that is, it is able to proceed by being able to take its origins and its fundamental task for granted; it owes its ongoing mode of being to its capacity to move away from and in a certain sense forget its origins. But in exchange for this very capacity to move forward, it always runs the risk of not only forgetting but also being unable to reactivate and critically examine its origins. And if the origins are faulty, the heirs to the tradition may inherit such faults through too little critical awareness of what they owe to the past.

After working on a translation of *The Crisis of European Sciences*,² in which Husserl developed at length his notion of the *Lebenswelt*, I am convinced that there are many faults and confusions in his exposition which need to be sorted out and examined. This task seems especially important since I suspect that some of Husserl's confusions have been handed down to his successors along with his profoundest insights. This latter point is something I shall not try to establish here. But

even the suspicion provides warrant enough for reopening the case, especially in this instance, since some have come to see the investigation of the life-world as either synonymous with phenomenology itself or as forming its most profound stratum. Husserl scholars point out, quite rightly, that the master himself did not see it this way, that he regarded this investigation of the *Lebenswelt* as merely a necessary preliminary stage on the way to transcendental subjectivity. "Original" phenomenologists reply, also quite rightly, that being true to Husserl is less important than being true to the "Sachen selbst," and it is precisely his move to transcendental idealism they object to. Nevertheless, they do credit Husserl with the discovery of the life-world and see themselves as continuing in the exploration of the terrain to which he led the way.

Husserl would have rejoiced in seeing himself thus characterized as a sort of Moses to the children of Thales, for he claimed the role for himself often enough. And he lived up to it all too well in one sense, as his readers know, at least in those works either published or meant for publication. He has a maddening tendency to describe in rough outline, as if discerned from the heights of Mount Nebo, the salient features of a new domain, confident that this will provide his successors with a reliable map with which to venture forth and fill in the details. More than any of his other books, the *Crisis* exhibits this character of outlining a program and issuing anticipatory directives. And this is true especially of the 90-page section devoted to the life-world. There is evidence, in fact, that this section was the very last thing to be inserted in the plan of the *Crisis*, forming an innovation even over the Prague lecture on which the work is based. It was apparently written during the very last year of Husserl's active life, a year of feverish activity interrupted again and again, and finally interrupted for

¹ Herbert Spiegelberg, *The Phenomenological Movement: A Historical Introduction* (The Hague, Nijhoff, 1960) vol. I, p. 159.

² *The Crisis of European Sciences and Transcendental Phenomenology. An Introduction to Phenomenological Philosophy*, soon to be published by Northwestern University Press. In the following I shall quote from my translation, referring to the pages of the German original (*Die Krisis der europäischen Wissenschaften und die transzendente Phänomenologie*, etc. [*Husserliana*, vol. 6], ed. by Walter Biemel (The Hague, Nijhoff, second printing, 1962)).

good, by illness. Husserl never claimed that it was more than a rough outline, of course, but as such it needs to be re-examined. For as explorers know, a faulty map of the terrain to be explored can cause grave difficulties to the exploration itself, setting it off on wrong paths and disorienting it from the start.

What I wish to argue in the following is that Husserl has assembled under one *title* a number of disparate and in some senses even incompatible *concepts*. Each of these concepts has some validity and importance, but Husserl does not seem to be fully aware of their separateness and thus does not concern himself with showing that or how they belong together. The question, then, is whether these notions can legitimately be combined under the title "life-world," and if so, whether the resulting clarified conception can play the role in phenomenology that Husserl thought it should play.

I

It is not actually in the *Crisis* that the term *Lebenswelt* makes its first appearance in Husserl's vocabulary; in fact, it appears in a manuscript meant as a supplementary text to *Ideas*, vol. II, dated by the Louvain archivists at 1917.³ It appears there closely related to a term that is familiar to readers of Heidegger and Merleau-Ponty: *natürlicher Weltbegriff*, natural world-concept,⁴ and it is linked to the investigations of part III of *Ideen II* concerning the construction of the personal, spiritual, or cultural world as opposed to the scientific or natural world. A comparison would reveal that many of the themes and descriptions of *Ideen II* and the *Crisis* are similar on this point, and that Husserl's later writings on the subject were thus able to draw on reflections initiated at a much earlier date. Nevertheless, it is only in the *Crisis* (1936) that Husserl self-consciously uses the term *Lebenswelt* with emphasis, employing it in the title of a major section of the projected book and according to the

elaboration of the notion a decisive position in the phenomenological program. After a brief introductory section on "The Crisis of the Sciences as Expression of the Radical Life-Crisis of European Humanity" (Part I) and a longer historical section devoted to "The Clarification of the Origin of the Modern Opposition between Physicalistic Objectivism and Transcendental Subjectivism" (Part II), Husserl turns to the lengthy third part, "The Clarification of the Transcendental Problem, and the Related Function of Psychology," which was to be the central section of the projected but unfinished five-part work.⁵ The first half of this third part bears the sub-heading "The Way into Phenomenological Transcendental Philosophy by Inquiring Back from the Pre-given Life-world," and it constitutes the longest single division of the book as it stands.

But by the time Husserl begins this section he has already prepared the way for the concept of the life-world; in fact the notion is introduced gradually in the framework of the historical discussions of Part II, beginning with the long section devoted to Galileo. If we look closely at the conception that emerges there, we shall be able to see clearly *one* of the several themes which are interwoven, as I claim rather confusedly, into the notion of the life-world later on.

According to Husserl, Galileo's great accomplishment, to which modern science owes its success, was the mathematization of nature. Husserl asks: "What is the meaning of this mathematization of nature?" and he re-phrases the question as: "How do we reconstruct the train of thought which motivated it?"⁶ This question sets the tone for the long, quasi-historical inquiry which follows, and Husserl's answer is pre-figured in the next paragraph.

"Prescientifically, in every-day sense experience, the world is given in a subjectively relative way. Each of us has his own appearances," Husserl says, and points out that these may be at variance with one another, a fact with which we are all familiar. "But we do not think," he goes on, "that, because

³ *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie, Zweites Buch* (Husserliana vol. 4) ed. by Marly Biemel (The Hague, Nijhoff, 1952), p. 375. See the "textkritische Anmerkungen," p. 423.

⁴ This term apparently derives from Richard Avenarius' book *Der menschliche Weltbegriff* (first published in 1891; third edition: Leipzig, O. R. Reisland, 1912) where it is the title of the first section. Heidegger implies that his *Sein und Zeit* (1927) provides the first adequate elaboration of this concept (eighth edition, Tübingen, Niemeyer, 1957, p. 52), but it was also discussed in detail by Husserl, not only in the manuscript mentioned but also, for example, in the *Phänomenologische Psychologie* lectures of 1925 (Husserliana vol. 9, ed. by Walter Biemel; the Hague, Nijhoff, 1962, p. 87). It is often claimed that the *Crisis* (1936) was influenced by Heidegger's book, and in some respects this is true. But it seems clear that Husserl's concern with the *natürlicher Weltbegriff* is at least as old as Heidegger's. See Merleau-Ponty, *Phénoménologie de la Perception* (Paris, Gallimard, 1945), p. i.

⁵ See Fink's "Outline for the Continuation of the *Crisis*," *Die Krisis*, pp. 514 ff.

⁶ *Die Krisis*, p. 20.

of this, there are many worlds. Necessarily, we believe in *the* world whose things only appear to us differently but are the same. [Now] have we nothing more than the empty, necessary idea of things which exist objectively in themselves? Is there not, in the appearances themselves, a content we must ascribe to true nature? Surely this includes everything which pure geometry, and in general the mathematics of the pure form of space-time, teaches us, with the self-evidence of absolute universal validity, about the pure shapes it can construct *idealiter*—and here I am describing,” Husserl notes, “without taking a position, what was ‘obvious’ to Galileo and motivated his thinking.”⁷

Here Husserl has described in a few words both the brilliant insight upon which modern science rests and the fateful mistake which has consistently misled the various attempts at its philosophical interpretation. Galileo inherits “pure geometry” from the Greeks as a science which affords exact, inter-subjectively valid knowledge for its domain of objects. In our encounters with the real world we have the problem of the subjective relativity of what appears, and it is the task of a science of the world to overcome this relativity. Now pure geometry is not unrelated to the world; in fact, as a science it can be seen as originally arising out of the practical needs of accurately surveying land and the like, and its theoretical formulation has always found application back to the real world. Galileo sees that this is because the real world as it presents itself to us in experience contains, somehow embedded in it, examples of what is dealt with so successfully in geometry. Galileo’s proposal is that exact and intersubjectively valid knowledge of the real world can be attained by treating *everything about this world* as an example of a geometrical object or relationship. If every physical shape, trajectory, vibration, etc., is seen, after being measured as accurately as possible, as a version of a pure geometrical shape, geometrical statements about the properties and relationships among these pure shapes will turn out to provide us with information about nature which shares in the exactness and universality of pure geometry. This leaves untouched, of course, certain properties which do not seem directly measurable in geometrical terms: color, warmth, weight, tone, smell, etc. Galileo notes, however, that changes in some of these properties correspond exactly to measurable changes in geometrical properties—even the Greeks had known of

the relationship between the pitch of a tone emitted by a vibrating string and its length, thickness, and tension. In his boldest move of all, Galileo proposes to treat all such “secondary qualities,” as they were later called, exclusively in terms of their measurable geometrical correlates with the idea that *all* will be accounted for thereby.

Thus is accomplished, according to Husserl, the mathematization of nature, and such is the origin of mathematical physics. It can be broken down into two steps, actually: Galileo’s *geometrization* of nature, and the *arithmetization* of geometry accomplished by Descartes and Leibniz. Nature becomes a mathematical manifold and mathematical techniques provide the key to its inner workings. In mathematics we have access to an infinite domain, and if nature is identified with that domain we have access not only to what lies beyond the scope of our immediate experience, but to everything that could *ever* be experienced in nature, i.e., to nature as an infinite domain.

It is by contrast to the Galilean conception of nature that Husserl’s first characterization of the life-world emerges. The philosophical interpretation of Galileo’s mathematization becomes involved in a series of equivocations. To overcome the vagueness and relativity of ordinary experience, science performs a set of abstractions and interpretations upon the world as it originally presents itself. First it focuses upon the shape-aspect of the world, to the exclusion of so-called secondary qualities; then it interprets these shapes as pure geometrical shapes in order to deal with them in geometrical terms. But it forgets that its first move is an abstraction *from* something and its second an interpretation *of* something. Its first move is an abstraction because, no matter how successful we may be in correlating secondary with primary qualities, the world we are trying to explain still presents itself to us as having both kinds of properties, one of which we systematically ignore or declare “merely subjective.” Its second move is an interpretation because, to treat the spatial relationships of the world with geometrical exactness, it must consider these relationships as the ideal ones with which pure geometry deals, whereas the real shape-aspect of the world, no matter how accurately measured, can never present us with anything but approximations to these ideal relationships.

Having forgotten the abstractive and idealizing

⁷ *Ibid.*, p. 20 f.

role of scientific thought, the philosophical interpretation comes up with an ontological claim: *to be is to be measurable* in ideal terms as a geometrically determined configuration. Thus it happens, says Husserl, "that we take for *true being* what is actually a *method*."⁸ Mathematical science is a method which considers the world *as if it were* exclusively a manifold of idealized shape-occurrences; the ontological interpretation simply states that it *is* such a manifold. The ontological claim then gives rise—and such is the course of modern philosophy—to a sequence of epistemological absurdities, the mathematical realism of the rationalists and the subjectivism and ultimately the scepticism of the empiricists. Rationalism treats the scientific method as if it were a kind of instrument, like the microscope, which allows us to *see* the world as it actually is, which pulls back the curtain of appearances and puts us into contact with reality. Empiricism recognizes that all we ever *see* is the causal effects of the real world upon the mind, and it raises the ultimately insoluble question of whether what we *see* accurately informs us of what *is*. The curtain of appearance is thus lowered again for good.

Husserl's critique is directed not so much against Galileo's methodical innovations as against those ontological and epistemological consequences drawn from it. The scientific method is not an instrument for improving our *sight*, something invented during the Renaissance which enables us once and for all to put aside the world of appearances. It was and remains an abstraction from and interpretation of what *is* seen, and what *is* seen remains ever the same whether or not we are scientists who operate with the method. This is the "world of sense experience,"⁹ the "intuitively given surrounding world [Umwelt],"¹⁰ as Husserl first calls it, or finally, the "prescientific life-world."¹¹ It is that *from* which science abstracts and *of* which it is the interpretation, the world of objects possessing both primary and secondary qualities, the world of spatial aspects belonging to vague and approximate types and not a world of geometrical idealities. On the other hand it is a *world* and not a mental representation of the world. It is "subjectively relative" by comparison to the intersubjective agreement the scientific interpretation affords, but it is not "merely subjective" in the sense that it belongs to the mind.¹² And most

important, the life-world is the "meaning fundament,"¹³ as Husserl says, of natural science, if natural science is correctly understood; for as an abstraction-interpretation, science would have no meaning, make no sense, without reference to that *of* which it is the abstraction-interpretation.

II

When he begins the section devoted directly to the notion of the life-world, Husserl picks up many of the themes that emerged from his critique of modern science and philosophy. Science operates with abstractions, the life-world is the concrete fullness from which this abstraction is derived; science constructs, the life-world provides the materials out of which the construction arises; the ideal character of scientific entities precludes their availability to sense intuition, while the life-world is the field of intuition itself, the "universe of what is intuitable in principle," the "realm of original self-evidence"¹⁴ to which the scientist must return to verify his theories. Science interprets and explains what is given, the life-world is the locus of all givenness. The emphasis here is on the *immediacy* of life-world experience in contrast to the mediated character of scientific entities. The life-world is prior to science, prior to theory, not only historically but also epistemologically, even after the advent and rich development of the scientific tradition in the West.

It must be said that in the context we have been describing, the *Crisis* offers us little that is new. Much of Husserl's actual description of the life-world at this point is simply a recapitulation of the phenomenology of perception with which readers of the *Ideas* and the *Cartesian Meditations* are familiar. The life-world is primarily a world of perceived "things," "bodies." He speaks of the perspectival character of perception, of outer and inner horizons, placing more emphasis than before, perhaps, on the role of the living body and its kinesthetic functions and on the oriented character of the field of perception around the body. His descriptions correspond to those centered around the concept of the "world of pure experience" in the *Phenomenological Psychology*,¹⁵ the analyses of passive synthesis and pre-predicative experience found in *Erfahrung und Urteil*.¹⁶ The critique of the distinc-

⁸ *Ibid.*, p. 52.

⁹ *Ibid.*, p. 21.

¹⁰ *Ibid.*, p. 22.

¹¹ *Ibid.*, p. 42.

¹² *Ibid.*, p. 127 f.

¹³ *Ibid.*, p. 48.

¹⁴ *Ibid.*, p. 130.

¹⁵ *Phänomenologische Psychologie*, pp. 55 ff.

¹⁶ *Erfahrung und Urteil. Untersuchungen zur Genealogie der Logik*, ed. by Ludwig Landgrebe, third edition (Hamburg, Claassen, 1964), pp. 73 ff.

tion between primary and secondary qualities, in which Husserl follows Berkeley, is of course not new, nor is his insistence on the ideal character of pure geometrical structures in opposition to the realities of the experienced world. Husserl's greatest innovation in this context, in fact, concerns not so much his characterization of the life-world as his assessment of the status of science. Mathematization is seen not merely as one interpretative way of dealing with the world, but as a historical phenomenon which involves an original establishment and a handed-down tradition. Galileo inherits the tradition of Greek geometry and combines it in a fruitful way with the need for a science of the world. His successors, in turn, take for granted his way of *interpreting* the world—which Husserl regards as a kind of methodological proposal or hypothesis¹⁷—and go on to make great discoveries and theoretical refinements. Philosophers, also taking for granted Galileo's proposal, absolutize it into an ontological claim which then makes experience and knowledge incomprehensible. It is to this historically determined, modern scientific *view* of the world that Husserl wishes to oppose the world as it really presents itself, the pre-scientific life-world in which we always *live* but to which our theoretical reflection has been blinded by our scientific prejudices. This historical characterization of scientific thought does reflect, by contrast, on the concept of the life-world, for it implies that the life-world is *not* historically relative phenomenon but the constant underlying ground of all such phenomena, the world from which the scientific interpretation takes its start and which it constantly presupposes.

III

It is against this background of explicit and implied characterizations of the life-world that many of Husserl's remarks appear puzzling and, in my view, point to a second notion of the life-world which differs radically from the first. Very early in the life-world section, attacking Kant for taking the world as the scientific world and ignoring the role of the life-world in scientific experience, Husserl writes: "Naturally, from the very start in the Kantian manner of posing questions, the everyday surrounding world of life is presupposed as existing—the surrounding world in which all of us

(even I who am now philosophizing) consciously have our existence; and here are also the sciences, as cultural facts in this world, with their scientists and theories."¹⁸ The sciences as theories, then, together with the scientists as creators of the theories, are part of the life-world. Again and again, but almost always in passing, Husserl refers to the sciences as cultural facts which belong, presumably along with other cultural facts, to the life-world. As they arise, he says, they "flow into"¹⁹ the life-world, "add themselves to its own composition,"²⁰ and enrich its content.

At first Husserl might seem to be involved in a flat contradiction here, since he previously distinguished the life-world from the world of science and now seems to be putting them back together. Husserl is aware of this seeming contradiction when he writes: "the concrete life-world, then, is the grounding soil [*der gründende Boden*] of the 'scientifically true' world and at the same time encompasses it in its own universal concreteness. How is this to be understood? How are we to do justice systematically—that is, with appropriate scientific discipline—to the all-encompassing, so paradoxically demanding manner of being of the life-world?"²¹ But Husserl seems to regard this particular paradox, at any rate, as being easily resolved. For it is not quite true that the scientific world and the life-world, previously distinguished with great care, are now being merged.

What Husserl is adding to the life-world is not the world *as described* by scientific theories but rather the scientific theories themselves; and when he refers to them in this way he always adds: "as cultural facts" or: "as spiritual (intellectual) accomplishments [*geistige Leistungen*]."²² "[Science's] theories," he writes, "the logical constructs, are of course not things in the life-world like stones, houses, or trees. They are logical wholes and logical parts made up of ultimate logical elements. . . . But this . . . ideality does not change in the least the fact that they are human formations, essentially related to human actualities and potentialities, and thus belong to this concrete unity of the life-world, whose concreteness thus extends further than that of 'things'."²³ There is a difference between engaging in science, i.e., interpreting the world according to its methods, and living in a cultural world of which science is a part. "If we cease being immersed in our scientific thinking," Husserl writes,

¹⁷ *Die Krisis*, p. 37 f.

²¹ *Ibid.*

¹⁸ *Ibid.*, p. 106 f.

²² *Ibid.*, p. 132.

¹⁹ *Ibid.*, p. 115.

²³ *Ibid.*, p. 132 f.

²⁰ *Ibid.*, p. 134.

"we become aware that we scientists are, after all, human beings and as such are among the components of the life-world which always exists for us, ever pre-given; and thus all of science is pulled, along with us, into the—merely 'subjective-relative'—life world."²⁴ Here Husserl has accomplished a brilliant reversal. The scientist sees himself as overcoming the relativity of our "merely subjective" pictures of the world by finding the objective world, the world as it really is. Husserl shows that the scientist can just as easily be seen, by a shift in perspective, as a man who himself has a particular sort of picture of the world, and that as such both he and his picture belong *within* the "real" world, which Husserl calls the life-world.

Now with this Husserl may have resolved one paradox about the life-world, but he has left us with another. For in describing the life-world as a cultural world which can contain scientific theories as well as stones, houses, and trees, Husserl has moved into what by his own account is a very different phenomenological domain. As Husserl says, scientific theories are not things, and, what counts most for the phenomenologist, they are not *given* as things are; they are not objects of perception, they are not given in perspective, they are not, strictly speaking, even spatio-temporal. And the same thing is obviously true of other elements of the cultural world: institutions, such as the state, the university, the church, the Bureau of Internal Revenue, do not stand before us simply as objects to be perceived; nor do works of literature, protest movements, the generation gap. Elaborate and many-leveled constitutive analyses must be devoted to these phenomena if this world is to be understood, as Husserl himself insisted in *Ideas*, vol. II; and above all the role of language in structuring both the community and its world must be appreciated. How is this to be squared with the "world of immediate experience?" This cultural world may indeed be described as pre-theoretical, in the sense that it does not need to number among

its constitutive elements a scientific theory of the world, much less the particular sort of mathematical-scientific theory developed in the modern West. But such terms as "pre-predicative," "immediate," "intuitively given" are clearly out of place. Least of all can the cultural world be described as historically and sociologically non-relative, that is, as something which does not change with the times and circumstances. How can the term "life-world" be used for such disparate concepts?²⁵

Husserl is not unaware of one aspect of the paradox just described and seems to think he has taken it into account: this is the historical and possibly sociological relativity of the life-world considered as cultural world. In spite of the "subjective relativity" of the life-world by contrast to the objective scientific world, Husserl writes, "normally, in our experience and in the social group united with us in the community of life, we arrive at 'secure' facts; within a certain range this occurs of its own accord, that is, undisturbed by any noticeable disagreement. . . . But when we are thrown into an alien social sphere, that of the Negroes in the Congo, Chinese peasants, etc., we discover that their truths, the facts that for them are fixed, generally verified or verifiable, are by no means the same as ours."²⁶ It is in this connection that Husserl often uses the term "life-world" in the plural, such that different historical periods and social groupings have different life-worlds. One way to overcome this "cultural relativity," of course, is to go the way of objective science itself, leaving the life-world behind to reach objective, i.e., mathematically determined truth. Husserl then asks if we are left with nothing else to say about the life-world other than that it is culturally relative. "But this embarrassment disappears immediately," he writes, "when we consider that the life-world does have, in all its relative features, a *general structure*. This general structure, to which everything that exists relatively is bound, is not itself relative. We

²⁴ *Ibid.*, p. 133.

²⁵ I cannot agree with Kockelmans' claim that there is "a perfect correspondence" between the "life-world" of the *Crisis* and the "world of immediate experience" in *Phänomenologische Psychologie*, and that the *Crisis* formulation is simply "more comprehensive and desirable" (Edmund Husserl's *Phenomenological Psychology. A Historico-Critical Study*, [Pittsburgh, Duquesne University Press, 1967, p. 288]). It is true that the 1925 lectures deal with "The appearance of *das Geistige* in the world of experience" (*Phänomenologische Psychologie*, p. 110) and even refer to "die Erfahrungswelt als Kulturwelt" at one point (p. 113). But Husserl is quite clear that cultural objects and even persons, though they are "perceived" in a broad sense (p. 115), are not given in sense experience strictly speaking. Thus he finally proposes a reduction to the world of (strictly) perceived "things": "Offenbar ist diese Dingwelt gegenüber der Kulturwelt das an sich Frühere. Kultur setzt Menschen und Tiere voraus, wie diese ihrerseits Körperlichkeit voraussetzen" (p. 119). Actually the *Psychologie* is more "desirable" since it contains many of the distinctions so badly needed in the *Crisis*.

²⁶ *Die Krisis*, p. 141.

can regard it in its generality and, with sufficient care, fix it once and for all in a way equally accessible to all."²⁷ This structure is what Husserl calls the *a priori* of the life-world, the essence shared by all particular life-worlds, whatever their content, which makes them what they are.

Now these considerations, I maintain, important as they are, do not dispel the discrepancy described earlier between life-world as cultural world and life-world as world of immediate experience. It is quite correct to speak of the different "worlds" of different peoples and historical periods, and it is also quite correct, in my opinion, to seek the general or *a priori* structures belonging to any such world purely as such. But we should be clear on the fact that, in undertaking the latter task, we are seeking the general structures of the *cultural* world and not necessarily of the world of immediate experience. Several differences between the two types of inquiries suggest themselves immediately. First, phenomenological analysis of the cultural world will have to deal, and in fact must deal primarily, with the constitution of precisely those cultural entities whose mode of givenness was contrasted earlier with that of the perceptual world. Its first subject of concern must be the ontological status of the community as such and the conditions of the possibility of such phenomena as institutions, political organizations, literature, religion, and mores, whatever particular forms they may take. This is the farthest thing from a phenomenology of perception. Second, since the "life-world" in this cultural sense can change historically, its phenomenology must deal with the eidetic structures of such change, the essential conditions of any and all cultural transformations. The phenomenology of perception, at least on Husserl's own account, need not concern itself with such transformations, since perceptual structures do not change. Finally, the investigation of the cultural world must appreciate the structuring role of language and the communication based on it, while the world of immediate experience, according to Husserl, is distinguished by being pre-linguistic or pre-predicative in character.

Now this is not to say that the phenomenology of the cultural world is totally *unrelated* to the phenomenology of the perceived or immediately experienced world. In fact, it is of the utmost im-

portance to show the dependence of the cultural world upon the perceived world for its constitution, and this again according to Husserl himself. The cultural community is not something perceived, like a thing or a body, but neither is it given to us independently of perceived bodies; we know the community because we perceive other persons as members, representatives, or authorities of the community and because we perceive physical objects such as tools and books, factories and monuments, as its artifacts and documents. But the cultural world is precisely *dependent* for its sense upon the perceived world and is not *identical* with it. It represents a higher and distinct level of constitution, just as, to go back to the first of the *Logical Investigations*, reading and understanding a sentence represents a higher level than simply perceiving the words as physical configurations on the page.²⁸ The former is *founded* upon the latter, as Husserl would say, but is by no means *reducible* to it. What is needed is a stratified constitutive analysis like the one in the fifth *Cartesian Meditation* leading from straightforward perception to the experience of persons and, from there, to the much more complicated constitution of the community.²⁹

But notice that we have now placed the cultural world in the same position, relative to the so-called world of immediate experience, as the scientifically constructed world of mathematical physics. That is, the cultural world is a domain of entities and structures whose givenness is mediated by and founded on the spatio-temporal world of perception. No less than the scientific world, the cultural world has its meaning-fundament in the world of perception as the domain through which its structures are always mediated, in which its truths are always directly "verified" in our experience. To be sure, the character of the mediation and the mode of being of the entities that make up the two "worlds" are quite different. It could be said that the two types of "mediated" experience focus on different aspects of the concrete world. Both are historical in that a coherent development and transformation of truths about the world is essential to both. But the character of the historical development is different; as Husserl points out, especially in these later writings, the historical development of science is *culmulative*, at least ideally; our concept of what is true does not simply change from

²⁷ *Ibid.*, p. 142.

²⁸ *Logische Untersuchungen*, vol. 2 (fifth printing, Tübingen, Niemeyer, 1968) pp. 61 ff.

²⁹ *Cartesianische Meditationen und Pariser Vorträge* (*Husserliana* vol. 1), ed. by Stephen Strasser (The Hague, Nijhoff, 2nd ed. 1963), pp. 149 ff.

one time to the next but grows in a constant progression, with each new stage building upon the ones before it. In spite of these differences, however, the parallels are obvious: surely our degree of removal from the gold crisis, for example, or the "Establishment," is as great as our degree of removal from the electron, and our access to these two sorts of entities is in many ways similar, in any case necessitating simple perception at some stage.

IV

The argument I have developed thus far points to a serious ambiguity in Husserl's notion of the life-world and to a resulting structural mistake regarding its position on the phenomenological map. He begins by distinguishing the world of post-Galilean mathematical science from the world of everyday life or life-world. He tries to show the priority of the life-world, the way in which the scientific world is dependent on the life-world for its sense. But the phenomena ranged by Husserl under the term "life-world" turn out, as we have seen, to fall into two distinct strata, one of which is indeed prior to the scientific domain (the "world of immediate experience") but the other of which seems to be on the *same* phenomenological level as the scientific domain, in spite of its differences—that is, in respect to its derivative or mediated character. From this perspective it is confusing at best to use a single name for the two different concepts I have been discussing.

But we might ask how Husserl was able to fall prey to this confusion. Or, to put the question in a more flattering way: Do the world of immediate experience and the cultural world have something in common, something of which Husserl was aware in placing them together under one term? This question can be answered affirmatively in a way which partly justifies Husserl's use of the term "life-world" even though it does not exonerate him from the error of using it in a confusing way. But it complicates matters further, supporting my statement earlier in this paper that Husserl's term "life-world" involves not just two but several different concepts at once. A brief examination of how this is so reveals the great multiplicity of interests and directions of inquiry which motivate Husserl in *The Crisis of European Sciences*.

Three considerations point to elements that are common to the two types of world we found in-

involved in Husserl's "life-world." First, we must remember that the touchstone of the *Crisis*, and the point to which it returns again and again, is modern mathematical science, and, in general, the problem of the theoretical science of nature. Now something which emerges, not so much from the *Crisis* as from the important short paper on the "Origin of Geometry,"³⁰ is Husserl's claim that theoretical science depends for its possibility not only on the world of immediate experience, the perceived world, but *also* on the cultural and linguistic community and *its* world. To exist and construct its mathematically determined world, science must have at its disposal not only a whole system of language but also a system of culture in which certain truths can be shared and taken for granted as a basis for continued work. The cultural world and the world of immediate experience, then, whatever their differences, are alike in constituting the preconditions for the existence of science. This means that the phenomenological stratification developed earlier must be somewhat revised. It is not as if the world of immediate experience made up a primary level supporting a secondary level which can take the form *either* of culture *or* of the scientific domain. Rather, the scientific level constitutes a *tertiary* stratum built on the second or cultural level. This does not invalidate our point about the important differences between the first two levels, but it does justify their sharing the designation "pre-scientific" in the sense that together they form the foundation for the mathematized world.

A second point in Husserl's favor centers around a term that is used repeatedly in connection with the life-world, namely "pre-given" (*vorgegeben*). The pre-given is what is there in advance, that which is taken for granted, which is passively received by consciousness and forms the background for its activity in relation to the world and itself. In keeping with his growing emphasis on the cultural world in his later writings, Husserl in the *Crisis* and other late texts sees it as contributing to what is always and necessarily pre-given to consciousness. It is not only the world of pure experience, the *a priori* of the life-world in *this* sense, that consciousness takes for granted in its dealings with the world; it is also the cultural world and whatever prejudices and interpretations may derive from it. In conscious life, man may be without scientific upbringing and thus lack the scientific

³⁰ This paper appears as one of the *Beilagen* to *Die Krisis*, pp. 365 ff.

interpretation of the world. But he is never, Husserl means to say, without culture, and thus never without some view of the world which goes beyond its immediate givenness to perception. Thus the cultural world, like the world of pure experience, is a necessary ground (*Boden*) of conscious life; it is pre-given not only for the theoretical activity of the scientist but for any activity whatever.

Finally, cultural world and perceived world are united in the very important conception of the pre-theoretical. This is a slightly different point from the first in this series, which pointed to the priority of the life-world over the scientific world; for in using the term "pre-theoretical" Husserl refers primarily to consciousness and the different sorts of attitudes it can assume. He had come to stress much more than in his earlier writings that conscious life is not exclusively, not even primarily, a quest for objective truths about the world which could be combined into a coherent world theory. The "natural" attitude in this later period is rather different from the "natural attitude" of the *Ideas* which, when closely examined, turns out to be the philosophical theory of naive realism. "Original natural life," as Husserl calls it in the Vienna Lecture,³¹ for example, is not theoretical at all, but rather practical. For consciousness at this level, the world is the domain of ends to be attained, projects to be carried out, materials to be used in carrying them out. It is not a mathematical manifold of entities to be known with theoretical exactness, but a pre-given horizon of the useful and the useless, the significant and the insignificant, the relevant and the irrelevant. There is no denying the Heideggerian flavor in these later considerations of Husserl, and the question of influence is properly raised. But in any case we can see that both the cultural world, with its instruments and socially determined projects, and the world of immediate experience must be seen as the *milieu* in which the pre-theoretical, practical life of consciousness runs

its course. In much of what Husserl says about the perceived world here, one is reminded of Merleau-Ponty's warning that perception must not be analyzed as if it were an "incipient science." The orientation of the perceived world around the lived body is a *practical* orientation of movement and accomplishment, not a theoretical orientation. Similarly, culture does not essentially present us with a "theory" of the world, but envelopes us in a domain articulated according to spheres of action, providing norms and directives for getting around. The cultural world may contain a scientific theory among its elements, but is not exhausted in the stock of objective truths the theory provides. Not that the concept of *truth* has no relevance here, for hand in hand with Husserl's new descriptions of consciousness and the world goes a new concept of truth. Here he refers to "situational" or "practical" truth,³² which is properly characterized as "merely relative"—i.e., relative to the subject or the community, relative to the project under consideration—only by contrast to the notion of "objective" truth, truth-in-itself about the world-in-itself.

The cultural and the perceived worlds combined, then, form the horizon of "natural" or primordial conscious life with its pre-theoretical attitude. And as such they form the pre-given ground from which the theoretical attitude arises, the pre-scientific world underlying the scientific. As I have said, these considerations go some distance toward clearing up the confusions built into the celebrated concept of the life-world and offer some justification for Husserl's rather broad use of the term. But at the same time they indicate that much more work needs to be done. Moreover, I think that if taken seriously they raise profound problems for the whole phenomenological enterprise, at least as its founder originally conceived it. In any case we should be warned that Husserl's concept of the life-world is not something that phenomenologists can simply take for granted.

Yale University

Received September 3, 1969

³¹ *Die Krisis*, p. 327. The Vienna Lecture ("Die Krisis des europäischen Menschentums und die Philosophie") in another of the *Beilagen*, pp. 314 ff.

³² *Die Krisis*, p. 135.

VII. REAL POSSIBILITY

BENJAMIN GIBBS

NO adequate metaphysic can do without the idea of possibility, yet explaining it is troublesome. This paper resolves a few problems only by ignoring or engendering others.

I shall try to distinguish some of the main ways in which we think and speak about possibility, hoping thereby to supply an engine for the destruction of bad arguments and a prophylactic against philosophical confusion. I shall mark distinctions by suitable adjectives qualifying "possibility" and its cognates. The definitions labeled by these terms are intended to elucidate ordinary discourse about possibility. I shall discuss primarily what I call *real* possibility, i.e., the kind which, unlike mere *formal* possibility, is correlative to *actuality*. Actuality as meant here is not something which can belong to mathematical objects or formal truths. It is the mode of existence proper to causal agents and patients. Consequently, all assertions of real possibility are tensed. I shall use also the idea of *incompatibility*. Two propositions are incompatible when the truth of each excludes that of the other, and two things or states of affairs¹ are incompatible when the actuality of each excludes or prevents that of the other.

There is one other preliminary point of importance. Sometimes, when asked whether something is possible, one is unsure whether to answer affirmatively or negatively, either answer being correct depending on various factors about which the question has not been explicit. The more obvious way in which this can happen is that something which is possible on one interpretation of "possible" may be not possible on a different interpretation. But also, any possibility has to be specified in terms of some description, which may be generic or highly specific, may refer to essential or to only incidental features, and so on. A relatively general description may specify a possible state of affairs while a more determinate description does not. This is not a point about the limitations of human knowledge. Whether or not unicorns may be said to be possible creatures depends less on epistemic factors than on whether we mean

by "unicorn" merely an animal rather like a small horned horse, or an animal like a small horse with a horn of pure gold and magical powers.

I. NATURAL POSSIBILITY

This is in a way the primary kind of real possibility, yet I am unable to define it adequately.

A thing is *naturally* possible if and only if nothing that already is or has been actual is incompatible with the actuality of that thing.

One might say this is only a nominal definition, if it were proper so to describe the explication of a concept which is restrictedly topic-neutral. None of the familiar attempts at explaining natural possibility, or explaining it away, seems convincing. The idea of *incompatibility* has to be explained using modal notions, and this means that the above definition is really circular. Incompatible things or states are such in virtue of their natures. Each is actual at a given time only if the other is not; and the "if" here is not a truth-functional "if." It is the "if" which occurs in both (a) "If sugar is put in water it will dissolve" and (b) "If sugar is wrapped in polythene and put in water it will not dissolve." To say that the fact that a lump of sugar is wrapped in polythene is incompatible with its dissolving in water is to say not merely that no polythene-enclosed sugar dissolves in water (when the water is on the other side of the polythene), but that the nature of polythene excludes or prevents water from dissolving polythene-enclosed sugar. If we employ the ideas of *exclusion* and *prevention* to explain that particular notion of incompatibility which is relevant to real possibility, we shall be using an unanalyzed idea of natural necessity or impossibility, which in turn will be nominally definable in terms of an unanalyzed idea of natural possibility. We may, however, use a notion which is characterized correctly, though inadequately, in the explanation and distinguishing of other notions. There are different kinds of incompatibility, but "incompatible" has a determinate and uniform

¹ The phrase "things or states of affairs" here is neutral with respect to actuality. Things are not *as such* actual.

meaning in all the definitions of kinds of real possibility.

The definition must be understood as presupposing that something or other is actual. Without this stipulation it would follow according to the definition that if nothing existed at all everything would be naturally possible. But the possible is a function of the actual. Nothing can become actual except by the operation or interaction of some other, antecedently actual agent or set of agents. It is true that, generally, facts about the present and past are respectable candidates for actuality. But if nothing existed at all, this fact would not have actuality but only truth. For all its consequences would be negative facts linked to it not causally but by entailment. I do not know how to construct a definition of natural possibility which would apply whether or not anything actually existed.

Particular varieties of natural possibility, such as physical possibility or chemical possibility, are distinguished from each other (and from the possibilities which are the subject-matter not of any natural science but of common knowledge simply) by the specific character of certain types of actuality. An assertion that such and such is chemically possible implies a reference to the basic elements and compounds of matter. For something to be chemically possible is for there to be nothing in the chemical nature of past or present actuality which would prevent the actuality of that thing.²

Philosophers who are attracted by the idea that the proper objects of scientific knowledge must be eternal and unchanging might argue that the sciences are concerned primarily with possibilities which are unrestricted in spatial or temporal location and which can be expressed in tenseless propositions. But if this were so, it would not follow that we should seek a tenseless definition of natural possibility in general. Many things are naturally possible at one time or place but not at another. It was not naturally possible in the eighteenth century, but is now, that one should visit London and Vienna on the same day. Particular physical possibilities are restricted in time and place; e.g., the possibility which obtains during a thunderstorm that certain crops will be damaged. Anyway, physical possibility is as much determined by the nature of past and present actualities as are the possibilities studied by biologists, engineers, or cooks.

The non-actuality of a present or past thing entails that the present or past actuality of that thing is now impossible. It would be a contradiction to say "It is possible that my name is 'Plato,' though actually it is not," for the second part of the sentence reports a present actuality which is incompatible with the actualization of what is mentioned in the first part. But some present-tensed possibility-assertions are true though the possibility be not actualized. The statement "It is possible that my name should have been 'Plato,' though actually it is not" is true. It says that at some time in the past there was nothing which was incompatible with my being called "Plato," though actually I was not.

Most assertions of natural possibility are tacitly or explicitly hedged with qualifications by their authors, with "unless . . ." clauses or *ceteris paribus*³ clauses. This is because there are hardly any true universally quantified propositions about what happens in the natural world. It is a general fact of nature that white shirts when put into blue dye turn blue; but if one were to put a white shirt which had previously been soaked in some special chemical into a vat containing blue dye, it might, as a result of the presence of the chemical, remain white, or turn green or some other color. The fact that white shirts when immersed in blue dye turn blue, unless the circumstances are unusual in specifiable ways, is not incompatible with a particular white shirt's turning green on being put into blue dye when the circumstances are unusual. So one might say it is a natural possibility that a white shirt when put into blue dye should turn green. Probably someone who denied this would not mean to assert unconditionally that it is impossible for a white shirt when put into blue dye to turn green. He would have been prepared to add ". . . unless the circumstances are unusual in ways *X*, *Y*, *Z*." If he intended such a qualification to be understood by his hearers, and if it would have been reasonable to expect them to recognize this intention, one could say that his assertion was true. For the fact that such and such a state of affairs never obtains unless conditions *X*, *Y*, or *Z* are fulfilled is incompatible with the actualization of that state of affairs in the absence of the fulfillment of any of those conditions.

The existence of tacit escape-clauses is a variable

² This is scarcely an adequate account of chemical possibility. A chemical theorist might be able to produce something more specific. Cf. W. E. Dasent, *Non-Existent Compounds* (New York, 1965).

³ No weight should be put on the phrase "*ceteris paribus*." We should probably get nowhere if we tried to explain *what* has to be equal when we say "other things being equal." Other phrases bring other difficulties.

element in assertions of natural possibility which is distinguishable from the variability in the description under which the alleged possibility is characterized. But the same elements may appear in the specification of the qualifications governing the assertion, or in the description under which the possibility is characterized. If one were asked to justify an assertion of natural possibility, it might make no odds whether one chose to do so by demonstrating the actuality of something mentioned in the escape-clause, or by showing that the terms in which the thing was described were such as to allow its possibility.

When something is actual, this is due to other actual things; to natural agents, not formal principles. It is a matter of geometrical necessity that the interior angles of any Euclidean triangle should add up to two right angles, and therefore that the sum of the interior angles of a particular triangle drawn with chalk should be two right angles; but it is not a matter of geometrical necessity that that actual triangle should exist. Again, it is a matter of semantic necessity that any bachelor be unmarried; but given some particular bachelor, considered as an actual person and not merely *qua* bachelor, it is not a matter of semantic necessity that he be unmarried. Any necessity involved will be natural necessity, deriving from such facts as that there is a dearth of suitable partners, or that he does not want to get married. Now it might be thought that the negative character of the definition of natural possibility will let in possibilities of a formal and therefore alien kind. It is arithmetically possible that the square of some number should be divisible by three, for the square of three or of any number which is a multiple of three, is divisible by three. The truth of this proposition is not excluded by any past or present actuality. Yet arithmetical possibility, not natural possibility, is what is involved. The reason is that a mathematical proposition is a candidate not for actuality but for truth. Any actual thing would instantiate various general facts of nature, but the fact that the square of some number is divisible by three does not.

The idea of natural possibility does not exclude what Quine calls "recalcitrant experiences"; nor does it exclude events which happen by magic, or miracles. Recalcitrant experiences are experiences which are recalcitrant to explanation in terms of our presently held theories. They are experiences of states of affairs which had been thought to be naturally impossible till they occurred, whereupon it became known that what had been taken to be

unrestrictedly general facts of nature were not facts at all, or were not unrestrictedly general. Similarly, if something happens by magic, though it may not be explicable in ordinary scientific terms it is not naturally impossible. A magical happening *could* be explained by someone acquainted with whatever principles might underlie the practice of the art of magic. Ordinary usage may collapse the concept of something which happens by magic into the concept of something recalcitrant to explanation, but strictly, the latter concept is more generic than the former. A magical event is not just a happening which is inexplicable by ordinary scientific principles. It is a happening which is explicable only in terms of occult or supernatural forces. The sources of magical power were supposed to be mysterious, mighty intelligences: gods or demons. The word "supernatural" may suggest that gods and demons do not have *any* nature. It means rather that their nature is not accessible to study by us using the ordinary means at our disposal. A magical event would be not an event above or against nature, but an extraordinary event resulting from the operation of some extraordinary agency.

The difference in the case of miracles lies only in the agency responsible for the event. Miracles are the work of God, maybe acting through human or other deputies. They are events which are naturally impossible except by God's agency. Thus it is naturally impossible that water should change into wine or that a man should rise from the dead. But there may be no natural impossibility about God's changing water into wine or raising a man from the dead. God has a nature which constitutes part of the subject-matter of theology. If God were to perform a miracle, one might be in difficulties about how to describe in detail exactly what he had done. Sometimes God's purpose might be achieved by conjuring up an appearance to his chosen witness without bringing about any real change. Moses' burning bush might have been a divinely-induced hallucination. But the change of water into wine at the marriage feast must have been a real change, for the drink had the essential properties of wine: it made the guests merry. It is no objection to the concept of a miracle that science could describe a miracle only at the level of nominal definition and could not explain how God had brought it about. God alone knows how he does it.

It is possible that there *should have been* dragons. At some time in the fairly distant past, nothing

that was then or had been actual would have prevented dragons coming into existence after that time, and still existing now. But because of how things are, it is not possible that there *should be* dragons now. This might be denied on the ground that God could suddenly create some. One might think of the author of all created nature as being such that a special escape-clause, "... unless God intervenes," has to be attached to every general proposition about the created world. On the other hand, frivolity and impulsiveness are not among the divine characteristics, and it is part of the idea of miracles that they are such rare events as not to be taken into account by natural scientists; and perhaps not even by lawyers.

II. STRICT POSSIBILITY

By the definition of natural possibility, if something is not and never has been actual, it is now not possible. Everything which is already actual is both naturally possible and naturally necessary. But maybe some future states of affairs are not yet naturally necessary. Maybe it is not the case that for every future state of affairs there is already something which prevents the non-actuality in the future of that state of affairs. A philosopher who believes that the future is not yet settled and unalterable in the way that past and present actualities are will not remain satisfied with the definition of natural possibility alone. He will require as well a definition such that things which cannot now be non-actual do not count as possible, while some future things may count as possible yet not at present naturally necessary. I call what such a philosopher requires, the concept of *strict* possibility. This is defined by adding to the definition of natural possibility the further stipulation that, at the time when the possibility obtains, its actualization be not yet determined.

A thing is *strictly* possible if and only if both

- (i) nothing that already is or has been actual is incompatible with the actuality of that thing, and
- (ii) nothing that already is or has been actual is incompatible with the non-actuality of that thing.

The definition presupposes that something or other is actual, and it implies that the fact that a future thing will never be actual does not entail that that thing is already now naturally impossible.

Strict possibilities are not the same as things which occur fortuitously or by chance. A thing which happens by chance is a coincidence of events or properties, such that independent sets of factors result, severally, in the occurrence of each event or property, but there is no *further*, independent set of factors which results in the coincidence of those events or properties. For example, it was a matter of chance that the man selected to design the University of Sussex had a beard. He was not chosen because of his beard. There is an explanation of why Sir Basil Spence was selected as the architect of Sussex University, but the explanation involves no reference to his beard. There is an explanation of why Spence was bearded, but the explanation involves no reference to his work as an architect. There is no additional explanation, over and above these, of why the architect of Sussex was a bearded man. The conjunction of the explanation of why Sir Basil Spence was selected to design Sussex University with the explanation of why Spence had a beard is sufficient to explain why the architect of Sussex had a beard. Now we can see why strict possibilities are different from things which happen by chance. At any time when a strict possibility obtains its actualization is not naturally necessary, but something which is naturally necessary may be a chance happening. It was a matter of chance that the architect of Sussex University was a bearded man, though at the time when this situation obtained it was naturally necessary. Conversely, some future things which at present are strictly possible will not, if they get actualized, do so by chance. It is now strictly possible that next weekend I shall go to visit friends in London. I have not yet decided to do so, nor not to do so, nor has anything else happened to settle things one way or the other. If I were actually to visit my friends, I could not be said to have done it by chance—though at present it is still possible I shall not visit them—for deliberate executions of decisions cannot be said to happen by chance. Only coincidences can be fortuitous.

It might be objected that every future thing is already determined by the nature of past and present actualities. The objection might be developed on fatalistic lines. With regard to any future thing either it is a fact that it will be actual or it is a fact that it will not be. Therefore it cannot be the case that both parts of the definition of strict possibility be satisfied. The mistake here is to treat facts about the future as actualities when they are only truths. What makes it true, if it be true, that I shall go to

London on Friday, is that I shall go to London on Friday. That future actuality bestows truth on "I shall go to London on Friday," but it does not bestow present actuality. The future actuality is not yet actual, nor is its actuality yet determined. Therefore the fact that I shall go to London is not a candidate for actuality.

Nevertheless, it might be argued, at any time the facts about what will happen subsequently are fully determined by the nature of what indisputably is or has been actual already. But how could this be established? It is not the case that, at some time in the past, the necessary and sufficient conditions of every subsequent event were already fulfilled. Many of the conditions of, say, my now having my present thoughts—that I should have been born and educated, and so on—were not fulfilled until comparatively recently. All that is certain is that, at some time in the past, all the conditions of my now having my present thoughts were fulfilled *except* the conditions which have come to be fulfilled since, and these latter have come to be fulfilled in consequence of the fulfillment of others traceable ultimately to the state of the world at that original time long ago. In short, every actual thing is a product of the fulfillment of some set of conditions jointly sufficient for its actualization, so that everything at the moment it becomes actual is naturally necessary. But we should not be forced to conclude that there are no strict possibilities until it were demonstrated that everything is naturally necessary at *any* time.

III. EPISTEMIC POSSIBILITY

Some philosophers use the phrase "empirical possibility"; but this is unfortunately ambiguous between natural possibility and epistemic possibility, and the ambiguity has led to confusion in many metaphysical and epistemological arguments.

A thing is *epistemically* possible if and only if there is no available way of verifying that there is something which is incompatible with the actuality of that thing.

For example: it is possible that at a certain time last week an albatross was flying over the North Atlantic at a certain precisely specifiable location. Let us suppose that there were no observers present, and now it is too late for there to be any point in sending observers, for the possibility relates to what happened last week. If actually there was no albatross there then, this would not entail that

it was not epistemically possible that there was one. *Proof* of non-actuality is incompatible with epistemic possibility; but not non-actuality *simpliciter*, so long as there is no available way of knowing about it.

It might be thought that epistemic possibility could be explained in terms of what *seems* to be possible, where "possible" has some non-epistemic interpretation. But something may seem naturally possible even when one knows it is not actual. And though a thousand years ago there was no available way of showing that men would never fly to the moon, it might have *seemed* impossible that they should ever do so.

Epistemic possibility is a commonly-used notion, but in one way the epistemic sense of phrases like "It is possible that . . ." is undeveloped in comparison with non-epistemic senses. It is true to say that three thousand years ago there was no available way of verifying that there was something incompatible with the earth's being a stationary body at the center of a cosmos of bodies in motion. But on its natural interpretation, the statement "It was once possible that the sun and stars revolved round the earth" is false. Epistemic possibility is not what is expressed in past-tensed possibility-assertions of this kind. The epistemic sense of the phrase "It is possible that . . ." is grammatically less developed than non-epistemic ones. This does not show that the epistemic notion is in some formal respect posterior to some non-epistemic notion. The idea can be used with reference to the past and the future, and applications of the definition can be tensed in all the standard ways.

There is a problem about the meaning of "available." It does *not* mean "available in principle," whatever that means. Perhaps though *I* do not know that there is something incompatible with the actuality of such and such a state of affairs, my neighbor does. It seems to be right to say "It's possible that Snooks was in the audience" so long as I do not know that he wasn't, though someone else—e.g., Snooks himself—knows that he was not in the audience. In this case the tacitly specified group of persons to whose knowledge the epistemic possibility is relative consists of just one member, the speaker. The flexibility of "available" allows "It is epistemically possible that . . ." to relate merely to the speaker's knowledge at the time (whether or not it would be reasonable to expect him to know more than he does, and whether or not there are relevant facts which he could easily find out if he took the trouble), so that it may mean no more than "I

don't know any reason why not. . . ." This flexibility in "available" does not entail homonymy in the term "epistemic possibility." There would be no point in distinguishing "objective" and "subjective" versions of epistemic possibility, for this would eliminate problems about "available" only if it were accompanied by the introduction of some such mysterious notion as that of "in principle."

G. E. Moore pointed out the fallacy which is sometimes committed by epistemological sceptics of moving from "It might have been that . . ." to "It may be that. . . ." A more accurate way of identifying the fallacy Moore wanted to expose would be to say that epistemological sceptics are prone to mix up natural and epistemic possibility. Mere linguistic clues, such as the use of the subjunctive mood or of the phrases "might have been" or "may be," are too unreliable. English has only the fossilized remains of a subjunctive mood, and deplorable though it may be, even some philosophers are so untutored as to be ready to accept "It is possible that my name is 'Plato'" as a way of saying "It is possible that my name should have been 'Plato'." The phrase "might have been . . ." often expresses natural possibility, and "may be . . ." epistemic possibility. But sometimes "might have been . . ." expresses epistemic possibility, as when one says "There might have been dragons long ago" meaning that there is no available way of verifying the opposite. The intuition of a difference between "might have been . . ." and "may be . . ." needs therefore to be supplemented by precise definitions, such as the ones I am in the course of putting forward.

IV. POSSIBILITY AND AGENCY

Natural possibility and strict possibility relate to the actuality of things and the obtaining of states of affairs. Possibility has to do also with the performance of actions, and with bringing it about that certain states of affairs obtain. The phrase ". . . is possible" may be elliptical for ". . . is possible for someone or something to do." The meaning of this may be further determinable in various ways. Sometimes "It is possible for A to ϕ " is equivalent to "It is possible that A should ϕ ," and expresses natural possibility. One might clear away a log which had been blocking a stream, and say "Now it is possible for the water to flow freely,"

meaning that there is now nothing to prevent it from doing so. But there are at least four other distinct notions which may underlie someone's saying that it is possible for an agent to do something. To define these we need the idea of what might be called the "inherent features" of a thing. These include its natural or essential features (those which determine what kind of thing it is, so that if it loses them it becomes something else), and those qualitative and quantitative features, including privations, which tend to persist in the thing notwithstanding changes in the external circumstances of its existence. The inherent features of a thing are to be contrasted with features of the situation or environment in which it happens to be at a given time, and with the relations the thing bears to those environmental or adventitious features.

V. POTENTIALITY

This notion covers the brittleness of glass, the power of a knife to cut, and so on. I shall call it "latent possibility."

An action is *latently* possible for an agent if and only if no inherent feature of the agent is incompatible with something's bringing it about that the agent does that action.

The analysis of *bringing it about that*, like the analysis of *causality* generally, has never been worked out satisfactorily by any philosopher. I take it for granted. The words "agent" and "action" here are not meant to be applied primarily to persons and their deeds. Note that the definition does *not* entail that the actualization of a latent possibility is brought about exclusively by something other than the inherent features of the agent in which the possibility resides. Winding a clock contributes to making it tick and move its hands, but so does the mechanism inside the clock. There is no inherent feature of a pane of glass which is incompatible with the pane's being shattered, though of course it will not actually shatter unless something makes it do so; unless the situation in which the glass is changes in some way (e.g., a missile strikes the glass) as a consequence of which the glass breaks. Latent possibility is therefore clearly distinguishable from tendencies, propensities, or dispositions,⁴ and also from natural possibility. It is latently possible for the glass in my bedroom window to shatter, but if I were to take special precautions to

⁴ No account of such ideas as *tendency*, *propensity*, *disposition*, or *inclination* is attempted in the present paper.

protect it there would be no natural possibility of its doing so. It is latently possible for a particular lump of sugar which is at present here in England to nourish the body of Mao Tse-Tung, for it has the potentiality to nourish the body of any person who consumes it. But it is not naturally possible that Mao Tse-Tung will eat it so long as he stays in China and the lump stays here, and therefore, though the sugar has the power to nourish him there is no natural possibility that it will.

VI. ABILITY

The notion of personal or quasi-personal ability or capacity I shall call "habilitive possibility," after the old adjective "habile." This is what is involved in "It is possible for A to ϕ " when this is equivalent to " A can ϕ " or " A knows how to ϕ " or " A has learned and not forgotten how to ϕ ." It might seem that this could be explained as a special case of latent possibility: but not so, for it might be that, though A is able to ϕ , he has some inherent feature which prevents anything or anyone getting him to do so. Snooks has the ability or capacity to commit blasphemy if he has learned a language and knows how to construct blasphemous expressions in that language. But he may be a pious person who will not commit blasphemy whatever happens. He has the ability to blaspheme but not the inclination or disposition. To make him disposed to blasphemy, for example by brainwashing, would be to *change* his inherent features. One could torture or trick him into saying something blasphemous; but then he would be held not really to have committed blasphemy, or not to have done so voluntarily. What we may correctly say about Snooks is that he has no inherent feature which is incompatible with his blaspheming (no lack of linguistic education, no vocal paralysis, etc.) unless it is that he has his will set against it. Or: the only inherent features of Snooks which would be incompatible with his blaspheming are ones incompatible with his not wanting not to do so. We may therefore define habilitive possibility as follows.

An action is *habilitively* possible for an agent if and only if any inherent feature of the agent which is incompatible with his doing that action is incompatible with his not wanting not to do it.

This account will stand whether the agent has either the opportunity or the desire to do the action, or both, or neither.

VII. OPPORTUNITY

"It is possible for A to ϕ " sometimes means " A has the opportunity to ϕ ." But "has the opportunity" is ambiguous. Usually having the opportunity to ϕ involves having the ability to ϕ , but sometimes not; though the potentiality to *acquire* the ability seems almost always to be presumed. It would be slightly odd for someone who does not know how to operate the controls of an aircraft to say "Last night I had the opportunity to steal an aircraft." Assuming that the only way of stealing an aircraft is to fly off in it, for someone really to have such an opportunity it must be the case that there is no obstacle at all to his acting. But opportunity does not always involve ability. One might be said to have an opportunity to speak to some foreign celebrity, but to be unable to take advantage of the opportunity because of not being able to speak the person's language. This is the less common meaning of opportunity. One would not say that a newly-born baby one was holding at the time also had the opportunity to speak to the foreign celebrity. Opportunity in the sense which does not imply ability I call "circumstantial possibility."

An action is *circumstantially* possible for an agent if and only if anything which is incompatible with the agent's doing that action is an inherent feature of the agent.

That is: there is no extrinsic obstacle to the agent's doing the thing, though he may have inherent features which exclude his doing it.

The more complex kind of opportunity, which is constituted by the combination of circumstantial possibility and habilitive possibility, I call "practical possibility."

An action is *practically* possible for an agent if and only if anything which is incompatible with the agent's doing that action is an inherent feature of the agent which is incompatible with his not wanting not to do it.

Practical possibility applies only to the actions of persons or quasi-persons for it involves habilitive possibility and with it the idea of *wanting*. There is no analogue to it in the case of inanimate bodies. The combination of circumstantial and latent possibility, assuming this to be permissible, would give not a new kind of possibility but actuality. The proposition "Anything which is incompatible with A 's ϕ -ing is an inherent feature of A , and no in-

herent feature of A is incompatible with something's bringing it about that $A \phi$'s" entails that nothing is incompatible with something's bringing it about that $A \phi$'s, which entails that $A \phi$'s. Anyway, it would be wrong to ascribe circumstantial possibility to inanimate bodies. It might sound all right to say that my lump of sugar has the power to nourish Mao Tse-Tung, but lacks the opportunity. On the other hand, it would sound queer to say that a tomato pip swallowed by Mao Tse-Tung had the opportunity to nourish him but lacked the capacity. If we do speak of "opportunity" in these cases we are not using the idea of circumstantial possibility, for it is not true that there is no extrinsic obstacle to the tomato pip's nourishing Mao Tse-Tung. There are features extrinsic to the pip but intrinsic to human digestive systems, which are incompatible with those systems' absorbing anything from tomato pips and therefore incompatible with the pips' nourishing them. The only ways in which one may take "It is possible for A to ϕ " when A designates an inanimate body are as expressing latent possibility or (when the sentence means "It is possible that A should ϕ ") natural possibility.

The differences between practical possibility and natural possibility help to explain why causal determinism does not exclude responsibility. By "causal determinism" I mean the principle that everything at the moment it becomes actual is naturally necessary. G. E. Moore said that in *one* sense of the word "could" nothing ever *could* have happened except what did happen. This is false if it means that there is some sense of "could" in which it would not be true to say that some things which have not happened could have happened. Perhaps Moore meant that nothing which has actually happened can now not have happened, in which case he was using the idea of natural possibility. He did, however, invent a technical sense of "could have done otherwise" to explain responsibility. The well-known objections to Moore's suggestion ("would have done otherwise, if . . .") do not undermine the plausibility of the idea that there is a link between responsibility and being able to do otherwise than one does. There is an application here for the idea of practical possibility. It is a necessary (though *not* sufficient) condition of someone's being held responsible for an act or omission that it should have been practically possible for him to do otherwise than he did; i.e., that he should have had both the ability and the opportunity to do otherwise; that the only thing preventing his doing otherwise was some inherent feature

of his which made him want not to do otherwise than he did. In this way, an action or omission may be free and the agent held to account for it, though at the time it happened it was naturally necessary. Some philosophers might think that absence of natural possibility of doing otherwise entails absence of ability, of opportunity, and of practical possibility. As the notions of *habilitive*, *circumstantial*, and *practical* possibility have been defined above, this is not so. It is sometimes said that, though a man may have the ability and the opportunity to act otherwise than he does, he may not be held responsible for his action if he may not be held responsible for his wants, which conflicted with his acting otherwise. I agree that many of a man's fixed wants or dispositions arise from chance factors or his education in such a way that he is not to be held responsible for them; but I deny that responsibility for a present action requires responsibility for all the past actualities which may have made that present action naturally necessary. What the sufficient conditions for responsibility are is however a question which lies outside the scope of my present enquiry.

VIII. REGULATIVE POSSIBILITY

There is another notion pertaining to what agents can or are able to do, but perhaps it should not be counted a kind of possibility. Nothing corresponding to it occurs among the meanings of "possible" listed in the standard dictionaries.

An action is *regulatively* possible for an agent if and only if it is not the case that there is some rule in force such that the agent's obedience to that rule is incompatible with his doing that action.

An example would be "It's impossible for us to drive down there," said with reference to a one-way street. The field of regulative possibility is the field of all actions which do not conflict with any normative principle, law, or rule. It covers moral and legal principles, the rules governing practical and productive activities, principles of technique, tactics, strategy, and so on. The commonest varieties are probably legal and moral possibility, though in common parlance we speak less of what is legally possible than of what is legally *permissible*, or of what legally one *can*, *may*, or *might* do. Regulative possibility is *not* connected with the so-called "laws of nature" which seventeenth-century deists believed in. It may be the case both that some

normative principle is in force and that people's actions do not conform to the principle; whereas it cannot be the case both that some natural fact obtains and that there is some actuality incompatible with it. As is often pointed out, expressions of rules are not statements of fact. With regulative possibility, incompatibility between *obeying* a principle and performing certain kinds of action is what is involved.

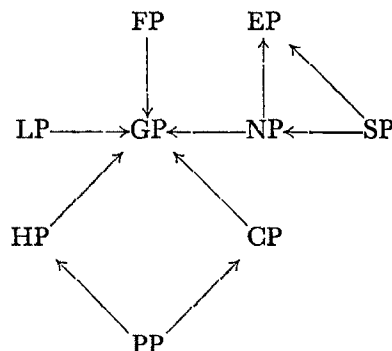
IX. GENERIC POSSIBILITY

Apart from regulative possibility, and epistemic possibility which is in another way *sui generis*, we have distinguished six kinds of real possibility. Each of the six definitions employs the idea of actuality essentially. In a way they are all members of a common genus, which they share with the various kinds of formal possibility. This genus is often called "logical possibility," but I prefer to reserve that label for something more specific, viz., the kind of formal possibility which is determined by the principles of formal logic.

A thing is *generically* possible if and only if describing that thing does not entail conjoining incompatible propositions.

The definition is not sufficiently general to cover epistemic or regulative possibility. But because of the many ways in which things can be incompatible, it covers not just real possibility but also the possibilities which are the subject of formal sciences such as geometry, arithmetic, formal logic, semantics, and metaphysics. These sciences are of course distinct. Incompatibilities which occur between propositions used in attempting to describe, e.g., bodies which are both red and blue all over are not due exclusively or even at all to truths of semantics, formal logic, or geometry. But I do not intend to attempt distinguishing kinds of formal possibility.

University of Sussex



The diagram illustrates the relations between the kinds of possibility, indicated by their initial letters. "FP" indicates "formal possibility." Note that, since epistemic possibility is not entailed by all the other kinds and does not entail any other kind, there is no *summu* genus of possibility.

For each of the definitions except those relating to agency, and most obviously in the case of epistemic possibility, we may construct a phrase to apply as a semantic operator to other kinds of possibility. It may be epistemically possible that something is naturally possible; or one might say "It's possible that it's possible for Snooks to dance," meaning that one knows of no reason why the circumstances in which Snooks now finds himself should prevent him from dancing. It would however be wrong to pretend that whenever people use the rare locution "It's possible that it's possible . . ." they have in mind something so complex or so precisely formulable as some of the examples we are now equipped to construct.

The distinctions made here would not be directly relevant to the interests of modal logicians, for a modal system requires only a single unambiguous interpretation of "possible" and a set of axioms which come out true for "possible" so understood. But the distinctions might have a rôle to play in a more complex, unified modal logic; no less elegant, and closer to ordinary ways of thinking, than the systems we already have.

Received October 20, 1969

VIII. ACTIONS AND EXTENSIONS

EVAN SIMPSON

MUCH current philosophical opinion endorses the view that human actions can be identified only by reference to the interests or intentions of the agent. Without such reference, some argue, it is impossible to distinguish what a person does from the consequences or implications of what he does, so that no action may have a unique description or be identified apart from its consequences or implications.¹ When one breaks a window (*A*) and thereby frightens the school-children (*B*) one obviously does not perform two different acts, but subsequently one may give two different answers to the question "What did you do?" This possibility warns against identification of *A* and *B*, even though they are not distinguishable deeds. That the boundaries of what a person does in a given situation are indistinct does not, however, imply that the boundaries of his actions are likewise indistinct, and arguments to the contrary seem to depend upon a momentary confusion about an otherwise untroublesome distinction between doing and acting. There may be two answers to the question "What did you do?", but ordinarily there will not be to the question "What act did you perform?". Only actions, not deeds, are governed by Leibniz' Law. It will be argued here that this is so, at least, for a perfectly definite class of actions, namely that comprised of actions which are neither necessarily intentional (like murder, marrying, signing a contract) nor necessarily intensional (like believing that such-and-such, seeing that such-and-such).² Deeds of these kinds may be described in importantly different ways according to the interests or beliefs of the agent or observer, and to avoid difficult problems about the identity, say, of writing one's name and signing a contract the class of actions under consideration may be restricted to exclude deeds sub-

ject in these ways to different constructions. This is simply to limit the present discussion to metaphysical aspects of the concept of action and to exclude those belonging to the theory of value or the philosophy of mind.

These restrictions make the thesis to be elaborated correspond closely to another current of philosophical opinion, according to which actions are as readily identifiable and uniquely describable as the agent who performs them.³ In particular, according to the present analysis action statements are extensional, equivalent terms being freely substitutable for one another; an apparent multiple polyadicity in the verbs of such statements is only apparent, terms the addition of which seems to alter polyadicity being expressible in statements conjoined to the basic action statement; because of this, desired entailments can be derived by simplification, existential generalization, etc.; attribution of intention and other normally adverbial features are expressible as operators on the statements in question; and although actions and events are spoken of as entities, no ontology of events is required in the sense that this is presupposed by the presence in a language of singular terms for such entities. In this last respect the view developed here differs from others recently formulated, maintaining that statements (but not terms) refer (but not singularly) to actions and states. This view thus differs, too, in the logical form attributed to statements of acts, reference to the agent, the object of action, and the tool employed supplanting need of any term for the action itself.

I

The states in question are taken to include states of affairs, processes, and events.⁴ It will be assumed

¹ Excellent statements of these positions are given by Eric D'Arcy in *Human Acts* (Oxford, 1963).

² P. T. Geach considers the spelling "intensional" an etymological atrocity. See *Reference and Generality* (Ithaca, 1962), p. 157 n. It nevertheless permits a useful, if largely unexamined, distinction.

³ This view is most strongly represented by Donald Davidson, "The Logical Form of Action Sentences," in N. Rescher (ed.), *The Logic of Decision and Action* (Pittsburgh, 1967), pp. 81-95.

⁴ The distinctions are G. H. von Wright's, *Norm and Action* (London, 1963), pp. 25 ff. The challenging task of clarifying intuitions regarding such states, if possible at all, has yet to be performed, and the characterizations of them here are necessarily undeveloped and tentative.

that such states may make contingent statements true, that is, that a statement is true when made in the appropriate state. The statement that *a* is ϕ 'ing *b* (that the bat is striking the ball, for example) is made true by *a*'s ϕ 'ing *b*, an event. Unfortunately, in spite of the close relationship between statements and the states in which they are true (a state-description being formed in the simplest cases merely by gerundizing the verb of the statement), the possibility of identifying states is not guaranteed by understanding the statements they make true. The required one-one relationship does not exist; one statement may be true in virtue of different states and a single state make various statements true. That is to say that statements tend to have a generic or a general quality foreign to the uniqueness of the states in which they are true. Thus the statements that *a* is ϕ 'ing something and that something is ϕ 'ing *b* can, because of their generality, both be made true by *a*'s ϕ 'ing *b*. The generic statement that *a* is ϕ 'ing could be made true by this or another state, and even the statement that *a* is ϕ 'ing *b* could be made true by any of a number of events of that description occurring in rapid succession, so that the generic quality imparted by the temporal indeterminacy of a statement in any tense could be eliminated only by construing the time in question to be a durationless instant within which nothing could occur.

To ignore this point may result in distinctly implausible constructions. Suppose that Jack fell down and broke his crown. Then, if times are taken to be discrete entities nicely ordered by conjunction, what is here said to have occurred is naturally construed as equivalent to the statement that Jack fell down *before* he broke his crown, which may in turn be expressed by "There exist times *t* and *t'* such that Jack fell down at *t*, Jack broke his crown at *t'* and *t* preceded *t'*."⁵ Yet the events in question cannot be so precisely localized in time. One cannot say of this case that, for example, Jack fell down *and then* he broke his crown (though "and then" serves the same function in this context as "before") because the temporal separation of events implied by this conjunction did not occur. One can easily convince oneself that there is no satisfactory way of specifying two discrete times, *t* and *t'*, in the above example and thus of the need for a fluid model of many temporal phenomena.

Because statements do not involve implicit reference to instants, it is safe to maintain that even

singular statements of states—whether past, present, or future—are generic rather than basic in the sense that they could in principle be made true only by a single state of affairs, event, or process, and the term "basic" may be reserved for another use, to be given it below. This conclusion amounts to the claim that true statements do not refer singularly to, or "name," the states in which they are true. The question whether such reference to states is possible at all is not mooted here, but it is a distinct advantage of the following analysis that the existence of a weaker relationship between statements and states is sufficient to establish the theses listed in the second paragraph of this paper.

Although the statement of an event cannot be used to identify the event which makes it true, it may reveal the structure and composition of the state or states in question. This is not to say that the composition and structure of a state may be discerned in any true statement, since this is obviously false for general statements, nor to say that at least the structure of a state may be discerned in any true singular statement, for were this the case no state would have a definite structure. That *a* is ϕ 'ing *b* implies that *a* is ϕ 'ing, so that *a*'s ϕ 'ing *b* and *a*'s ϕ 'ing cannot be considered distinct states. One need not conclude, however, that the only genuine state is that corresponding to a statement which implies all others describing the universe at a "given time" nor that states are not uniquely identifiable entities at all,⁶ for a simple expedient escapes extremity or indeterminateness. When a statement has all and only the name places its sense demands and these places are filled by names or their referential equivalents, reference is omitted to none of the entities which in relation make up the state in which the statement is true. The demands of sense here are simply a function of a verb's grammatical transitivity or intransitivity, so that although the statements that *a* is ϕ 'ing, that *a* is ϕ 'ing something, and that *a* is ϕ 'ing *b* may all be made true by the same state, the event which makes the last true making the first two true also, only the last meets the demands of sense. What is significant about this is that only the last could not be made true by states having different structures or constituents; and in virtue of its *referential completeness* it is made true only by states of a certain basic kind. Any state, then, which makes the statement true may be called *basic* in the sense that only it or an identically constituted state could

⁵ Thus, Donald Davidson, "Causal Relations," *Journal of Philosophy*, vol. 64 (1967), pp. 695-696.

⁶ Compare C. I. Lewis, *Analysis of Knowledge and Valuation* (La Salle, Ill., 1946), pp. 48-55.

make the statement true. It is useful here to introduce the concept of "assertive reference" such that a statement refers assertively to any state which makes it true, and in general a state may be said to be basic if it can be assertively referred to by a complete statement. By extension, complete statements may be termed "basic" as well.

"Complete" means "minimally complete." The criterion for referential completeness may be taken to prohibit variable polyadicity in verbs of complete statements.⁷ The statement, for example, that *a* is striking *b* with *c* at place *h* at time *t* because of *r* in the manner *m* . . . , is no more complete than the statement that *a* is striking *b*. In general, statements of states are complete when they have the number of references the verb demands, and when complete they need not be elaborated in ways not so demanded. "The bat is striking the ball" is not further completed by "in the stadium" or "with force *g*." "The cat is sleeping" need not be supplemented by "on the mat" nor "The sun is shining" by "on me." Grammatically reflexive verbs present interesting complications which will be considered briefly below. With an important exception characteristic of statements of certain actions, the only terms needed for completing a statement are the subject and the direct object or the predicate adjective. The first is always required and either the second or the third may be required in addition. Prepositional objects are never required.

It may be mentioned for future reference that the completion requirements for various verbs may be used to distinguish the several kinds of states mentioned above, that is processes, states of affairs, and events. Expressions for processes and states of affairs take neither kind of object, and in ordinary English statements of them typically have the form ϕx , thus, "It is raining," "The apple is red." The formally unrecognized grammatical distinction between participial and adjectival constructions seems to be without special significance in this context. It is interesting to note, however, that statements of processes and states of affairs indicate that the former may be identical to or degenerate into the latter. "The cat is sleeping" means the same as "The cat is asleep," and "He is dying" sooner or later gives way to "He is dead." Similarly, if an apple is reddening it will be that the apple is red, and one would at least be understood if instead of "It is raining" he were to say "It is raining." It may be noted, too, that states of affairs and

processes are caused but not causes. States of affairs, being static, may not cause other states, though they may be terms in other sorts of consequence-relation: The apple's being red may have as a consequence my eating it. Processes, likewise, can hardly be causes, given their identity with states or tendency simply to degenerate into them. Events, in contrast, can obviously be causes, and statements of them are formally distinguishable from statements of other sorts of simple state. Taking direct objects in ordinary English, they have the logical form ϕxy , "The bat struck the ball."

In brief summary of the last few paragraphs, the completeness of a statement has to do not with what may be expressed in a single utterance but with the minimal requirements for assertive reference to one or more identically constituted states. Additional references are superfluous because they contribute nothing toward the reference of a statement to such states. Additional references, therefore, must merely set out concomitants of a state, and although these may be produced at great length—if *a* is ϕ 'ing *b* it is doing so at a certain time in a certain way and so on (perhaps indefinitely)—such extended description of a situation does nothing to alter the reference of any statement to a given state. The description may be of great informative value for someone who is not acquainted with the state in question, that is, a person who does not know what state is referred to by the minimally complete statement, but it does not extend or restrict its reference to that state, any more than a description of an object, *a*, while useful to someone who does not know what "*a*" refers to, alters that designation. Additional references provide additional identifying conditions for a state, but they are not part of its identity conditions.

It is important that only terms or expressions in the name places of a statement are crucial to its assertive reference. Nothing precludes generic statements, even those expressed in a past or future tense, from being referentially complete and so to refer only to identically constituted states. That *a* ϕ 'ed *b* is true if ever a state describable as *a*'s ϕ 'ing *b* occurred, and there may be many such states, each of which is referred to by the statement. Since, however, the states are all relevantly similar, the basic character of the statement is undisturbed. A different sort of predicate inspecificity is as readily handled. The active voice is canonical for statements of events and statements in the passive must

⁷ This is to take account of the problem posed by Anthony Kenny, *Action, Emotion, and Will* (London, 1963), pp. 156–162.

be so translated to include the logical subject and to preclude the occurrence of prepositional object phrases.

If formal features of a verb cannot alter the completeness of a statement neither can the meaning of the term, so long as the referring expressions remain unchanged. Non-synonymous statements may refer to the same state, and the statements that the bat is striking the ball and that the bat is damaging the ball, for example, may be made true by the same event. That substitution of one verb for another may not alter the reference of a statement is subject, however, to the objection that truth-preserving substitution can change a statement about a cause into one about an effect. The bat's striking the ball occurs, say, at t , but since the bat's damaging the ball clearly involves an effect of the former event, that effect (the ball's being damaged) must be considered to come into being at a distinguishably later time $t + t'$. The difficulty is unreal because there is no such time. There is a sense in which cause and effect are indeed distinguishable in this situation, but only in a way that makes the effect be in a very strong sense similar to the cause: Cause and effect are related as event to end state, the ball's being damaged being an outcome of the bat's hitting the ball (=the bat's damaging the ball), and the distinction between the two *events* must be one of description and not of reference. In general two events are identical if they have the same constituents and if the expressions for them can be properly connected by "and" but not by "and then." In short, verbal inspecificity is usually irrelevant to assertive reference to states.

This concludes the outline of the method for determining the structure of basic states by means of referentially complete statements. The method does not provide a means for identifying individual states, but it does obviate certain problems about the nature of states, such as the status of "states" corresponding to incomplete statements and the question of whether any criterion for the composition of states will require that there can be only one state. Regarding "incomplete states," the relationship of assertive reference between statements and states is such that even if a statement may for want of a reference be incomplete, the states referred to by it cannot be. When a statement fails to refer to a specific state or states, it does not refer to an incomplete state but may refer to many different kinds of states. On the other hand, the whole world need not be taken to constitute the only state, since

it is possible to construct complete statements which do not involve reference to everything in the world. A statement which did refer to everything would be no more complete than a minimally complete one. Elaboration by means of additional references does not allow a statement to refer to any kind of state more specifically; instead it allows a statement to refer to more than one kind of state, and in this sense such elaboration increases specificity. By addition of a prepositional phrase another state may be referred to, namely that obtaining between the entity referred to by the subject of the statement and the entity referred to in the prepositional object phrase. Such statements, of course, owe their truth to more than one state and indeed to any or every state indicated.

II

For the simple cases examined to this point a statement containing n prepositional object phrases can be expressed as a conjunction of $n + 1$ relevantly related statements—i.e., statements each of which contain at least one term which is also found in another, and each of which express a state of affairs, event, or process. Statements of acts, though identical in grammatical form to the statements of events, etc., already mentioned, resist this kind of treatment. This is not immediately obvious, since many statements of action seem to accord with it. Consider the statement that Brutus killed Caesar in the Forum and notice that it does not state whether it was Brutus or Caesar or both who were in the Forum at the time. It is compatible with murder at a distance. Assuming that at least Caesar was in the Forum at the time, the statement of action may be represented by

$$(\exists x)(\exists t)(Kbc(t) \& Icx(t))$$

where K means "kills," I means "is in," t is a time and x is a place, and variables indicate features which are mere adjuncts of the act. Excluding the possibility of murder at long distance this is equivalent to

$$(\exists x)(\exists t)(Kbc(t) \& Ibx(t)).$$

Given this restriction, both propositions seem adequately to represent the complete statement of action, since if Brutus killed Caesar at t and Caesar (or Brutus) was somewhere at that time, there is nowhere else that Brutus' killing Caesar could have occurred then. The scene of the deed need not be specified in a specification of the action, and it

could be determined empirically. Being in the Forum was, therefore, no part of the murder but only incidental to it. The same considerations hold for the time of the incident, so that it need not be explicitly referred to in a complete statement of action. Consequently, the above statement may be expressed, in an exactly analogous way to event-statements containing prepositional objects, as "Brutus killed Caesar and Brutus was in the Forum." The only difference between this statement and one containing a temporal quantifier is that the conjunction is not purely truth-functional, since it expresses implicitly understood and not always explicitly expressible temporal relationships. From here on, therefore, symbolic renditions of statements will employ a non-truth-functional ampersand in place of the truth-functional dot. The above statement may then be represented more economically as

$$(\exists x)(Kbc \& Ib x).$$

The presence of a variable in one conjunct now indicates that no states making that conjunct true are part of the act in question. The absence of a temporal reference requires a slightly different interpretation. That Brutus killed Caesar on the Ides of March cannot be expressed; in accordance with the rules set down, either by "Brutus killed Caesar and it was the Ides of March" (there is no repetition of a referring term) or by "Brutus killed Caesar and Caesar died on the Ides of March" (there is a prepositional phrase). In no statement of the form "Such-and-such happened at t " can " $at\ t$ " be eliminated in a prescribed way. This is only to be expected. Everything happens at some time, and specification of that time is required for identification of the state or act involved. It is irrelevant only so far as the completeness of statements is concerned.

Although time and place need not be specified in complete statements of acts, there is another sort of reference which must be made, so that acts, unlike events, cannot be construed as simple two-termed relations. Although apparently similar to the statement that Brutus killed Caesar in the Forum, the statement that Brutus killed Caesar with a knife does not submit to such analysis. The similar compound

$$(\exists x)(Kbc \& Ub x).$$

meaning, "Brutus killed Caesar and Brutus used a knife," will not do, for it does not exclude the possibility that Brutus killed him with a club while using the knife for another purpose, nor that he

used more than one knife. It will not do, either, to say,

$$(\exists x)(Kbc \& Kxc)$$

meaning, "Brutus killed Caesar and a knife killed Caesar," for two persons may have killed Caesar jointly, Brutus using his club and an accomplice wielding a knife, and the possibility that Brutus used two knives remains. A satisfactory representation of the statement may be obtained only by combining the two previous propositions and instantiating the variable. One may then write

$$Kbc \& Ubk \& Kkc,$$

where k is the knife employed. Possible worries about ambiguity of "killed" are dispelled below. That Brutus killed Caesar, that he used a certain knife the while, and that the knife killed Caesar cannot be made true by any states having constituents other than Brutus, Caesar, and the knife in the relations indicated; the statement is, therefore, complete. The irrelevance of other states to the truth of the statement is guaranteed only by specification of the tool employed in the act, the action-verb "is killing" being a three-place predicate of a certain kind. This result shows incidentally, that the statement that Brutus killed Caesar in the Forum is incomplete. It should read, "Brutus killed Caesar with a knife in the Forum," which may be written

$$(Kbc \& Ubk \& Kkc) \& Ib f.$$

Brackets indicate the basic action statement and are otherwise irrelevant, simplification occurring as usual.

In a great many cases, then, an act may be expressed in the form

$$\phi xy \& \theta xz \& \phi zy$$

where x , y , and z are subject, object, and tool respectively, ϕ is an action and θ is use. In order to make explicit the paronymous relationship of the two occurrences of ϕ , one may be marked with a prime and the form of act-statements amended to

$$\phi' xy \& \theta xz \& \phi zy,$$

but because it proves possible to eliminate this encumbrance the nicety need not be scrupulously observed. Concomitants of the action may as before be expressed as clauses conjoined to the basic statement and adverbial modifiers attached as operators:

$$A(. .) \& -$$

No attempt will be made to provide rules governing the scope of ' A '; different adverbs may apply to

different conjuncts. In any case, however, the result of dropping it from a true statement is another true statement.

This general formulation excludes many putative acts. It ignores, for instance, cases in which the agent of an act or the tool employed are pairs. Brutus and Cassius may have killed Caesar jointly, or one may have used two knives. Such possibilities create complications, but they can be accommodated by the method and will not be considered here. More serious is the exclusion of statements which summarize series of actions. John may have eaten the tomato with the fork, for example, but it will hardly do to say that John ate the tomato, he used the fork, and the fork ate the tomato. What happened was that John speared the tomato with the fork, John raised the tomato with the fork, etc. "Build," "telephone," "fight," and a host of other verbs similarly express other than single basic acts, and the statements they form must be broken down into ones referring to such acts before the methods suggested here are employed.

Whereas the more complex things which people do fail to fit the act-schema, simpler ones can be accommodated. Although such basic bodily actions as moving one's arm or flexing one's muscles do not readily find place in the present scheme, place can be made by expressing such actions in degenerate act-statements. (In what follows the word "action" will label anything done by an agent which can be fitted into the act-schema; the word "act" is reserved for non-degenerate actions.) Events may similarly be presented as degenerate actions. "The bat struck the ball," for example, may be expanded to read, "The bat struck the ball, the bat used the bat, and the bat struck the ball": $\phi ab \ \& \ \theta aa \ \& \ \phi ab$. In events subject and "tool" are one and the subject "uses" itself, and the notion of self-use may be taken without loss to be the same as self-identity. After eliminating redundant and trivial conjuncts, then, the statement of an event reduces to ϕab . Little ingenuity is required to see how statements of processes and states of affairs may also be characterized in degenerate versions of the act-schema. "The sun is shining," for example, may surely be expressed tediously as, "The sun shines itself, it uses itself, and it shines itself": $\phi aa \ \& \ \theta aa \ \& \ \phi aa = \phi aa = \phi a$. These observations permit significant generalization.

There are four plausible cases of the expression " $\phi xy \ \& \ \theta xz \ \& \ \phi zy$," namely those in which

- (1) $x = y, x = z, x = y$, so that $\phi xx \ \& \ \theta xx \ \& \ \phi xx$

- (2) $x \neq y, x = z, z \neq y$, so that $\phi xy \ \& \ \theta xx \ \& \ \phi xy$

- (3) $x = y, x \neq z, z \neq y$, so that $\phi xx \ \& \ \theta xz \ \& \ \phi xx$

- (4) $x \neq y, x \neq z, z \neq y$, so that $\phi xy \ \& \ \theta xz \ \& \ \phi zy$.

There is a fifth consistent combination, namely

- (5) $x \neq y, x \neq z, z = y$, so that $\phi xy \ \& \ \theta xy \ \& \ \phi yy$

but the only available interpretation of such a statement places its possible referents outside the range of phenomena known to exist. It may be that one person could kill another by using him to kill himself, but, unless taken to mean simply that one killed him by getting him to kill himself, such an act would apparently require a direct exercise of psychical force. Lacking evidence for the existence of such an agency, only (1)-(4) are of present interest. Each corresponds to a familiar category of action, some examples of which may be given, explained, and finally interpreted. Typical of the forms are (1) "The apple is red," "The sun is shining," "Peter loves himself," "He is yawning." (2) "The bat struck the ball," "God created the earth," "Peter flexes his muscles," "Peter raised his arm." (3) "She tortures herself with those small shoes," "He killed himself with that knife," "The mouse fed itself with its paws." (4) "Peter shaves himself with a razor," "The dog scratched the door with its claws," "Peter raised his arm (left) with his hand (right)."

The principles determining these classifications are simple. If the verb in question is reflexive the statement must ordinarily be of the first or the third form: this follows from the identity of subject and object in (1) and (3) above. If in such cases another term is required for completion, then the statement must be of the third type, otherwise of the first: this, too, follows from the patterns of identity and difference characteristic of (1) and (3). Intransitive verbs may be considered trivially reflexive. If on the other hand a statement's verb is irreflexive, then it must be of the second or fourth type. The distinction between these two types is obvious, but one curiosity should be acknowledged. "Peter shaves himself" is grammatically reflexive yet appears in a place which demands it be irreflexive. Clearly this placement is proper; one does not shave one's self, but one's face or other part of the body. Notice also the difference between "Peter raised his arm" and "Peter raised his arm with his hand." Ordinarily one does not raise his arm with anything, but if it should be severed or paralyzed that might be necessary.

In general it seems possible to characterize the

actions expressed by statements of the four forms in this way: Those of category (1) are primitive mental states or involuntary processes; though actions in some sense, they are not within the scope of phenomena subject to the will or products of agent-causality. (Activities, like running, the statements of which have this form may be dismissed on other grounds as complex.) The actions of category (2) are basic bodily actions. Characteristically, such actions lack a distinguishable causal component, and the only answer possible to the question, "How did you do that?" is of the sort "I just did." The actions of category (4), acts, are the basic ways in which one deals with the world, but those of category (3) are, although quite like acts, rather odd. There is something peculiar about action directed toward oneself as the formalizations of the various kinds of action-statement tend to confirm.

In each of (1)–(4), the last two conjuncts may be taken to explicate the first, so that the first conjunct of each such statement is actually superfluous. It has been implicit in the preceding discussion that one may define the action of an agent in terms of the cognate event concept and the concept of use:

$$\phi'xyz = df \theta xz \ \& \ \phi zy.^8$$

In the first and second cases this explication is trivial, but it is not in the third and fourth cases. To say "Peter killed himself with a knife" ($\phi'xyz$) is just to say "Peter used the knife and the knife killed Peter" ($\theta xz \ \& \ \phi zx$), and to say "Peter shaved his face with the razor" ($\phi'xyz$) is just to say "Peter used the razor and the razor shaved Peter's face" ($\theta xz \ \& \ \phi zy$). In neither case does the explication suggest that Peter does something to himself, though in the former something certainly happens to Peter. This would seem to correspond to the common belief that suicide is ordinarily a sign of insanity, that one cannot will it with one's whole self, as it were. More generally, the reflexive actions of the third category seem to constitute behavior rather than acts, to be characteristic of sorts of personality or mental states, to manifest drives rather than to reflect particular goals. By contrast actions of the fourth category, acts, may (but need not) be explicitly teleological, or intentional.

Intentional or not, actions of every type can be

referred to and identified apart from any reference to the specific mental condition of the agent, his beliefs, interests, intentions, or whatever. That this is so is obviously the case for involuntary processes (yawning) and basic bodily actions (raising the arm), and it is arguable that self-love and the like are not exceptions in this respect. The contention is really problematical only for those actions the analysis of which involves reference to the agent's use of some tool. It may seem that the concept of use cannot be analyzed extensionally or that the use of a tool cannot be identified apart from the purpose for which it is used. To make the import of this objection clearer and to prepare the way for a reply, consider an interpretation of the concept of use which deprives that concept of any suggestion of purpose. Let it express a purely physical link with the agent, so that to use a knife, for example, is simply to hold or flourish it. If, then, "Peter killed himself with a knife" means "Peter used a knife and the knife killed him," that is just to say that Peter had a knife in hand which killed him. If Peter puts a knife through his heart, therefore, one may properly say that he killed himself, whether what happened was accident or suicide. The latter question has nothing to do with the identity of the phenomenon nor with its proper categorization. The deficiency of the analysis is obvious, however. It provides no way of distinguishing Peter's plunging the knife into his breast from someone's hitting the hand which holds the knife, thus causing a fatal wound. In the latter case it would not be proper to say that Peter killed himself. It follows that the concept of use, and thus an action of the third or fourth type, cannot be analyzed apart from some such notion as purpose. In some sense Peter must have meant to be doing with that knife what he was doing.

One can also conclude, however, that no reference to Peter's mental state is required to identify his action. The ability to refer to Peter as agent already presupposes all that is required. Because Peter can be distinguished from whatever might push his hand, the action in question may be identified, Peter's use of the knife distinguished from what is not such a use, without any knowledge of what is going on in Peter's mind. Given that one can tell Peter from a lower animal one can tell his actions from other phenomena, and only extensional criteria are needed for the identification of

⁸ This apparently supports the claim of D. G. Brown that the concept of personal action derives conceptually from that of physical or inanimate action. See his *Action* (London, 1968).

the phenomenon in question.⁹ Perhaps it would be better to call such criteria merely quasi-extensional, since they presuppose reference to persons, or agents, as entities of a certain kind, and are, therefore, not purely referential. Because the difference between reference and description is known to be imprecise, however, such scruples may obscure more than they clarify.

It is of interest to observe that the concept of use which remains incorporates the dictum "No action at a distance." It seems that this requirement cannot be relaxed. What would usually be expressed by "Peter killed himself with a gun," for example, is not plausibly analyzed as "Peter used a gun and the gun killed Peter." That suggests that he used the thing as a bludgeon, and the more likely deed is better expressed, "Peter used a gun and a bullet killed him." This sentence, however, does not express a basic act, and any satisfactory account of the situation will distinguish two or more distinct events. Consider another case. Suppose that some-

one broke the window with a rock and that this may be interpreted to mean either that he broke it by striking it with a rock or that he broke it by throwing a rock through it. Each of these involves specification of a subsidiary act, the person's striking the window with a rock in the first case and his throwing a rock (presumably with his arm) in the second. In the latter instance, specification of the tool shows that the rock is not the tool in the subordinate act, so that whereas in the first instance the subordinate act is evidently identical (contingently) to the act to which it is apparently subsidiary, two acts are distinguishable in the second case. This marks a distinction between tools and instruments and it may be concluded that, since a tool is a literal extension of one's body, projectiles are not tools. Deeds requiring analysis in terms of the broader concept of instrumentality cannot be immediately accommodated within the action-schema and require somewhat more elaborate treatment.

McMaster University

Received September 3, 1969

⁹ This point is made in a quite different way by Virgil C. Aldrich, "On Seeing Bodily Movements as Actions," *American Philosophical Quarterly*, vol. 4 (1967), pp. 222-230. It considerably lessens the interest of the view that the concept of personal action derives from that of physical action, since it provides no basis for thinking that the concept of an agent derives from that of an object.

IX. DOES BECOMING ENTAIL A CONTRADICTION?

JOHN KNOX, JR.

DOES time pass? Do events become? Do temporal facts have a transitory aspect? One need not answer "Yes" to these questions; in fact, an affirmative reply may turn out to be contradictory. But one who does say "Yes" is likely to wonder if the future has the completely determinate character of the present and of the past. Such a person may also have worries about the degree of reality possessed by past or future events, at times at which they *are* past or future. Most clearly and most importantly, his answer commits him to holding that events and temporal positions undergo a unique variety of change, called "passage" or "becoming": they are less and less future; then they are present; finally they are more and more past. By contrast, the thinker who responds with a "No" will deny that the events we call "past," "present," and "future" differ from each other ontologically. He will deny, in fact, that pastness, presentness, and futurity are characteristics. On his view, to attribute to an event what seems to be one of these characteristics is to do no more than describe the event as being earlier or later than, or as being simultaneous with, one's own utterance or some other particular event. Finally, the person who denies becoming has no sympathy for claims about purely temporal flow, movement, or change. An event can go somewhere in time no more literally than a road can go somewhere in space.¹ To deny becoming is not to deny that events or moments (i.e., temporal positions, such as 10:00 A.M., E.S.T. on June 22, 1985) are strung out in a distinctively temporal series, and that they exist successively rather than all at once. But it is to deny that events or moments change by having, at various moments, the characteristics of presentness and, in varying degrees, of pastness and of futurity.

I shall try to show in what follows that temporal becoming is self-contradictory and cannot, therefore, be a fact. No doubt such an enterprise sounds

naive—perhaps not even healthy. Yet in principle it is legitimate. For a description which entails a contradiction cannot possibly apply to anything in the world. I shall not undertake to decide if becoming is a necessary feature of events in time. But if it is such a feature, then I suppose one will have to side with McTaggart in denying that time itself exists or is real.² It will not be out of place, therefore, if I conclude by suggesting that a denial of time's reality is not totally meaningless, and may even have to be taken seriously.

Almost anyone would admit that becoming is highly paradoxical. For becoming is a kind of change; and change, one would think, presupposes time, and hence cannot be a factor within it. The difficulty becomes especially acute when one asks at what rate the change takes place. For to this question no possible answer seems informative, or even intelligible. It is true that becoming is no motion in space, and it is not change of the qualitative sort. Still, it is change; and change of any sort may presuppose time, and may have to take place slowly or quickly, smoothly or jerkily. One may reply, however, that although ordinary change—change *in* time—presupposes time, temporal becoming—change *of* time—does not; that, indeed the ordinary sort of change presupposes temporal becoming. And one may hold that the notion of rate applies nontrivially not to all changes, but only to those which, unlike temporal becoming, are subject to some sort of measurement. Since we cannot conceive of a device with which to measure becoming, the notion of a rate of change has no meaningful nontrivial application to it.³ Thus, the difficulties cited so far may establish nothing more exciting than that the change which constitute becoming is *sui generis*—a far cry from the conclusion that it does not exist.

The impression of paradox remains, however despite such reassurances. And as I shall now try to

¹ Cf. Donald Williams, "The Myth of Passage" in Richard Gale, *The Philosophy of Time* (New York, 1967), p. 105.

² See J. M. E. McTaggart, *The Nature of Existence* (Cambridge, 1927), vol. II, chap. 33, pp. 9–31.

³ Here I am following the approach taken by Richard Gale in *The Language of Time* (New York, 1968), pp. 241–243.

show, becoming conceals, not merely a paradox, but an actual contradiction.

Prima facie, there are two ways in which becoming may be viewed. One may choose to regard it as a movement of moments or of events from future to present to past. Or, equally well, one may see it as a movement of pastness, presentness, and futurity along a series of moments or of events stretching from earlier to later.⁴ Yet what appear to be two descriptions are really one. For purged of metaphorical associations, both amount to the fact that a series of temporal characteristics—what McTaggart calls the *A*-series—and a series of moments or of events ordered by relations of precedence and subsequence—what McTaggart calls the *B*-series—change relatively to each other in such a way that a given event or moment will be present, say, at one moment rather than at another. Now if at a certain time one event is more past than a second, or is present, the other being future, or is less future, then the one event is earlier than the second, the second later than the first. Thus, the fact of becoming, if it is a fact, entails that certain events are earlier, or later, than others; indeed, in being applied to all events it entails all relations of precedence and of subsequence, as well as of simultaneity. I shall express this by saying that the application, at a given time, of the temporal characteristics of the *A*-series to any given set of events, including the set of all events, entails the relative positions of these events on the *B*-series. Thus, if for events to become, and thus to be subject to *A*-series characteristics, should be inconsistent with their having the *B*-series positions which their *A*-characteristics at a given time entail that they have, then becoming would be inconsistent with itself. And similarly, if for events to become should be inconsistent with their having any *B*-series positions at all, then becoming would be inconsistent with itself; for if given events have no *B*-series positions at all, they certainly do not have the *B*-series positions which their *A*-characteristics entail that they have.

Relations of subsequence and of precedence are unchanging, as is the relation of simultaneity. Thus it is impossible for *M* to be earlier than *N* at one time, but later than *N* at another. In fact, it is contradictory to speak of relations of earlier and

later, or of simultaneity, as obtaining at certain times rather than at others. (The relation of precedence, for example, spans the interval between different times. How, then, could it obtain at either time?) What is true for singular events is true also for classes of events in relation to each other or to given events, or to events occurring at given times. Thus, it is contradictory to ask *when* (some or all) gold rushes succeeded the Norman Conquest, or preceded the events of the year 2000 A.D. For even if certain gold rushes did succeed the Norman Conquest, it cannot be true that they did so at one time rather than at another. In a similar way it is contradictory to assert that in 1066 many skirmishes preceded the discovery of America.⁵ The above facts constitute part of what I may call "the logic of *B*-series relations." I shall assume that if becoming should turn out to be inconsistent with the logic of *B*-series relations, then becoming would be inconsistent with itself. For we have seen that in application to given events, the concept of becoming entails their relative positions on the *B* series, and in turn that they *have B-series positions and relations*.

If becoming is a fact, then the feature of presentness changes its *B*-series position with respect to a given event, say the writing of this paper. Thus, let the letter *M* stand for that event. Then two months ago presentness characterized only certain events earlier than *M*. At the present moment, however, presentness characterizes only events simultaneous with *M*, including *M* itself. And a year hence, presentness will characterize only certain events later than *M*. Now the characterization of an event by presentness must be simultaneous with the event itself. Thus, if my son was born at 12:24 A.M. on March 3, 1969, then presentness must have characterized this event at precisely that time; and *vice versa*. It follows, therefore, that the characterization by presentness of an event which is earlier than *M* must itself be earlier than *M*. Thus we have:

(1) All characterizations by presentness of events earlier than *M* are (tenselessly) events earlier than *M*.

But we know that two months ago, say on March 1, presentness characterized only certain events earlier than *M*—or, in other words, that on

⁴ Cf. McTaggart, *op. cit.*, pp. 10–11.

⁵ Very possibly these questions and utterances are sheer nonsense. But any nonsense which exists is due ultimately to the fact that one element in the question or utterance entails or presupposes something which is denied by another such element. So even if the question or the utterance is nonsense, it is contradictory as well.

March 1 the only characterizations of events by presentness took place in connection with events earlier than *M*.⁶ And so we may write:

(2) On March 1, all characterizations by presentness were characterizations by presentness of events earlier than *M*.

Note that (2) follows directly and unavoidably from the assumption of becoming. For the assertion of a change is equivalent to the assertion of successive states or states of affairs; and (2) is the assertion of one among many such states of affairs which collectively constitute becoming.

One should see that (2) cannot be analyzed or translated by

(2') All characterizations by presentness occurring on March 1 were characterizations by presentness of events earlier than *M*.

For whereas (2) conveys the existence on March 1 of a complex state of affairs, (2') associates with March 1, not this total condition, but only all characterizations by presentness. Since an event and its presentness occur together, the entire state of affairs must, indeed, exist or occur on March 1. But what (2) asserts, (2') merely implies, and that only with the help of the further premiss that an event and its presentness have the same temporal boundaries. The difference will be plainer if we take a pair of statements about things or objects—e.g.:

- (a) In 1925, all autos were extremely quiet.
- (b) All autos on the road in 1925 were extremely quiet.

If someone makes the statement (a), it would be wrong for me to respond, "But you haven't said *when* they were quiet. Was this in 1925, in 1926, or perhaps many years later?" Such a response would, however, be appropriate should someone make the statement (b). Granted that all autos running in 1925 were quiet, it has *not* yet been said *when* they were quiet. Thus one might complete (b) by the words, "by 1940—when the last of them went to the junk yard." In (2), therefore, the force of "On March 1" is to make all that follows relative, as a whole, to this day. For this reason, (2) is a legitimate answer to the question, itself legitimate if becoming is a fact, "When were all characterizations by presentness such characterizations of events earlier than *M*?" We may grant that (2') could be substituted for (2), giving us a different statement con-

sonant with the hypothesis of becoming. Yet this hypothesis does, also, justify us in asserting (2)—and we choose to do so. And (2') is not a translation, let alone an articulating analysis, of that assertion.

Now as untensed, (1) is true without regard to temporal restrictions. We may therefore treat "characterizations by presentness of events earlier than *M*" as a middle term, and conclude:

(3) On March 1, all characterizations by presentness were events earlier than *M*.

A comparable argument would be the following:

Scott is (tenselessly) the author of *Ivanhoe*.

At *T*, the only person left was Scott.

Therefore, at *T* the only person left was the author of *Ivanhoe*.

Observe that this conclusion is placed in jeopardy only if we are hesitant about treating the major premiss as tenseless, perhaps wondering whether Scott could be the author of *Ivanhoe* before he wrote it. Note, further, that should (1) be trivially analytic we should have no fear of substituting "events earlier than *M*" for "characterizations by presentness of events earlier than *M*" in (2). Yet the typical analytic statement is tenseless. Thus the fact that (1) is tenseless is no ground for suspecting the role of (1) as a premiss in the argument, above, for (3).

Similarly, the characterization by presentness of an event simultaneous with *M* must itself be simultaneous with *M*. Thus,

(4) All characterizations by presentness of events simultaneous with *M* are (tenselessly) events simultaneous with *M*.

But just now, on May 1, presentness characterizes only those events simultaneous with *M*. That is,

(5) At present, on May 1, all characterizations by presentness are characterizations by presentness of events simultaneous with *M*.

And from (4) and (5) we may conclude:

(6) At present, on May 1, all characterizations by presentness are events simultaneous with *M*.

I shall not go through the steps, but parity of reasoning requires us to accept:

(7) A year hence, all characterizations by presentness will be events later than *M*.

Now the characterization of an event by presentness is itself a certain sort of event. Our result,

⁶ I am assuming here, and I shall assume hereafter, that all events are of fairly brief duration, say a few days at the maximum. This assumption will not compromise the argument, since I could have chosen to refer to all characterizations by presentness of events having a certain duration, rather than simply to all characterizations by presentness.

therefore, is that at a certain time, all such events were events earlier than *M*; that at another time—the present—all such events are events simultaneous with *M*; and that at still another time, all such events will be events later than *M*.

But each of the three parts of this result contradicts the logic of *B*-series relations. If the members of a class of events are earlier than (later than, simultaneous with) a certain specified event, then they stand in this relation irrespective of times. It is possible to say, "The California gold rush of 1849 preceded the elections of 1968." And it is possible to say, "In 1875, gold rushes were numerous." But it is *not* possible (consistently or intelligibly) to say, "In 1875, all (some) gold rushes were events earlier than the elections of 1968." For if gold rushes were such earlier events, they were not such events at one time rather than at another. Now characterizations by presentness are, certainly, peculiar items. And yet, they are not too peculiar to be events; and as such they must obey the logic of *B*-series relations. I conclude that becoming entails a contradiction. It must, therefore, have the status either of a misinterpretation of temporal experience, or of an illusion embedded in that experience.

The difficulty may be interpreted as follows. If we accept becoming, then we must grant that *B*-series relations connect events or moments on the *B*-series with characteristics on the *A*-series, and do so in different ways at different times. For this reason, we are entitled to speak of the *B*-series relations which present events, or less ambiguously events of presentness, have at certain times rather than at others. Yet relations of precedence, of subsequence, and of simultaneity are such that they cannot obtain *at* any given time at all; and for that reason they cannot change. The problem, then, is that on the hypothesis of becoming, *B*-series relations must serve to relate, not only events and moments on the *B*-series, but also the *B*-series with the *A*-series. And this second function proves self-contradictory. We may grant that the terms of the *A*-series are not events, but characteristics; and this fact complicates the picture, requiring that we refer to the members of certain classes of events rather than to the particular events themselves. But as I have tried to show, the contradiction does remain, even when this fact about the *A*-series is taken into account.

If we give up becoming, then the sort of difficulty just encountered does not arise. It will still be the case that the events of March 1 are earlier than *M*, that the events of May 1 are simultaneous with *M*,

and that the events of the succeeding May 1 are later than *M*. But the second premiss in each of the above arguments will not be assertible. It makes no sense to say, "On March 1, all events were earlier than *M*"; and no prior hypothesis entitles one to say it. Here is where the defender of becoming finds his difficulty. For in his view events change with respect to their pastness, their presentness, and their futurity. A given event will be held to be present at, but only at, a certain time; and then one will be driven to make the statement that this event changes in its *B*-series relations to the class of present events, and has a particular relation to the members of this class at one time rather than at another.

At this point one feels like objecting that presentness is, after all, nothing else than existence, and so does not define a special class of events. But presentness is not, for the theory of becoming, the same thing as existence. If it were, then those who defend becoming would be claiming nothing not granted by those who deny it. Everyone agrees that events exist or occur successively. What some affirm and others deny is that events have the temporal features of pastness, of presentness, and of futurity, and change with respect to these features. If presentness is merely existence or being, and is thus not a characteristic of any kind, then pastness and futurity are not characteristics either. If they were, then presentness would be at least the characteristic of standing temporally between pastness and futurity. So if presentness and existence are the same, then pastness and futurity are nothing else than existence before or after some given event. And in that case there is nothing left of becoming. We must affirm, therefore, that for the doctrine of becoming, presentness is a characteristic of events.

But is an event's being present to be considered an event in its own right? Here one question may be answered by another: Can it be considered anything else, when the change constituted by becoming is a change in its own right, and is not just the original event itself? Either something happens to the event, or it does not. If it does, then its happening surely is an event, and an event other than the event to which it happens. But if nothing happens to the event after all—if the event merely exists or takes place at a certain time—then the event does not change with respect to its temporal characteristics, and becoming has been denied.

Another point may be questioned. I spoke in my argument of the date of an event *M*. But perhaps I had no right to do this; for it may appear that what

has a definite date is not *M*, but *M*'s presentness. Yet on any theory which accepts the reality of time, *M* must be granted a date. If we did allocate *M*'s would-be date to the event of *M*'s being present, then by the same principle we should have to reallocate the date of *M*'s presentness to the presentness of *M*'s being present, and so on forever.

A passage from a recent book by Richard Gale suggests a linguistic formulation of the main argument of this paper, and it will be well if I indicate in outline what such a formulation would be like. Gale writes as follows:

My next utterance of "now" will denote a different time, even if I just wait where I am, but my next utterance of "here" will not denote a different place unless I move about. This difference between here and present or now is due to the fact that there is no spatial analogue to temporal becoming: the present (now), unlike here, shifts inexorably, independently of what we do.⁷

It would appear, then, that the successful use of the word "now"—or, as we may say equally well, of the present tense—may be substituted for presentness. Following this clue, we arrive at the following course of reasoning.

The only time at which the present tense may be used to describe an event is the very moment of that event. Thus, any present-tense-descriptions of events earlier than *M* (the writing of this paper) must themselves be earlier than *M*:

(8) All present-tense descriptions of events earlier than *M* are (tenselessly) events earlier than *M*.

But we know that two months ago, on March 1, the present tense could be used to describe only certain events earlier than *M*—or, in other words, that the only present-tense-descriptions were of events earlier than *M*:

(9) On March 1, all present-tense-descriptions were present-tense-descriptions of events earlier than *M*.

It follows that on March 1, all descriptions of events by means of the present tense were earlier than *M*:

(10) On March 1, all present-tense-descriptions were events earlier than *M*.

Similar reasoning will show that

(11) At present, on May 1, all present-tense-descriptions are events simultaneous with *M*;

and that

(12) a year hence, all present-tense-descriptions will be events later than *M*.

Now (10), (11), and (12) are severally contradictory. The assumption, therefore, that the successful use of the present tense shifts with time leads to a contradiction, and must be rejected. And in that case it is evident that becoming must be rejected as well. This form of the argument may be stronger, in that it does not treat presentness as a characteristic, or an event's being present as itself an event. It would appear, though, that the hypothesis of becoming is appealed to in the present form of the argument as much as in the first, and that it is this hypothesis which is responsible for there being a similar result in the two cases.

To return briefly to the argument in its first form, one begins by pointing out that if becoming is a fact, then presentness changes its *B*-series position relative to *M*: it starts out by being earlier than *M*; then it is simultaneous with *M*; finally it is later than *M*. One is thus allowed to say that on a certain date, say on March 1, presentness belonged only to events earlier than *M*. But the characterization of an event by presentness must be simultaneous with the event itself. On March 1, therefore, all characterizations of events by presentness were earlier than *M*. Yet this statement violates the logical principle that if the members of a certain class of events bear a *B*-series relation to a given event, they do not do so at any particular time. Now the fact that events have *B*-series relations follows from the assumption that events and times become. That assumption thus is inconsistent with itself, and must be given up.

The question whether time itself must be rejected will not be considered in this paper. But I should like to ask in conclusion what, if anything, a rejection of time's reality would mean, and how plausible such a rejection would be. What is said here will of necessity be brief and, I am afraid, somewhat dogmatic.

A denial of time's reality could only mean, first of all, that there are no such determinations as past, present, and future, and that nothing happens earlier than or simultaneously with anything else. Such a claim might be objected to on linguistic grounds. Aren't temporal expressions used to describe certain plainly contrasting states of affairs? And if so, can a denial of time and of

⁷ *The Language of Time*, p. 214.

temporal distinctions be anything else than a recommendation that we refuse a use to certain expressions for which we now have a use, indeed a perfectly good one? But such an argument confuses use in the context of meaning and communication with use in the context of application. No one denies that I can usefully go on meaning something different by "two hours" from what I mean by "one hour." But the question is whether either meaning applies to anything in the world. Similarly, "elf" means something different from "witch." But I can easily wonder whether either term succeeds in describing the world. To be sure, "two hours" and "one hour," unlike "elf" and "witch," are today very useful expressions. Yet errors need not be haphazard; and a systematic error could be quite useful within its own system, even though the entire system were wrong.

Again, one may hold that temporal expressions have been ostensively defined. And how, in that case, could they possibly fail to have application? To know what a temporal expression means or how it is used is already to have encountered an instance of the sort of thing to which the expression applies. But such an argument confuses an experience with the supposed object or event experienced. No one doubts that we experience the world as being in time; yet this fact is enough by itself to account for the possibility of ostensive definition of temporal terms. Ostensive definition requires, certainly, that one have a certain kind of experience. But it cannot be held, apart from further argument, to require the existence of the item defined; and for this reason it is incorrect to define ostensive definition in terms of an encounter with that item. For me to "get the idea" it is enough that I *seem* to experience the sort of item in question. Thus a successful ostensive definition need not be an encounter at all; it need only be a certain type of experience.

More important, I believe, than such linguistic difficulties is the problem: Is a denial of time's reality at all plausible, given that our experience suggests so forcibly that time is a fact? McTaggart attempts to render such a denial less paradoxical through an appeal to the specious present.⁸ Yet Broad's criticisms of McTaggart's effort seem effective, and I shall not concern myself here with the specious present or with McTaggart's attempted use of it.⁹ A more promising line of thought may

derive from McTaggart's comparison of our experience of time to the illusion of a bent stick and, again, to the appearance of things seen through red glass.¹⁰ These comparisons serve to make it plain that we are dealing with an erroneous or misleading experience rather than with a mere mistake in judgment. Yet as McTaggart points out, they fall short of providing us with adequate analogies. Unveridical sense perception relates only to spatial objects; the illusion of time covers our experiences as well. The sensory experiences just cited are particular and temporary; the illusion of time pervades all of our awareness. Those sensory illusions are easily identified as such by appeal to further sense experiences; a change of perspective will not enable us to view reality as it might be in itself, and apart from time. Can we, then, locate some feature of experience which, although fundamental and quite pervasive, we yet hesitate to ascribe to the reality experienced?

Depending upon the conditions of perception, whatever I experience through the senses I experience as appearing one way rather than another; and I have no way of comparing the object as it appears with the object as it might be in itself, and apart from any appearance at all. Here, then, we have a gap in our knowledge which mere experience cannot bridge; and what exists on the far side of this gap *may*, we recognize, be very different from what exists on the near. In themselves, objects do not appear. (To appear is to appear *to*.) As experienced, objects can only appear. Appearances, therefore, are pervasive elements or aspects of our sense experience which yet may, for all we *know*, be fundamentally misleading concerning the nature of reality.

Our sense experience provides us, then, with the beginnings of an adequate analogy. That it gives us more than that I cannot claim. It is one thing to hold that the ways things appear may be fundamentally misleading. It is a further step to claim that the way an item appears is nothing in its own right, indeed nothing real at all. Now if time is contradictory, we shall not be going far enough by condemning it to the world of experience. Instead, we must banish it totally. And whether the analogy with sense experience stands up at this point is certainly debatable.

If this paper is right, then we may side with logic

⁸ *Op. cit.*, pp. 27-30.

⁹ Broad's discussion appears in his *Examination of McTaggart's Philosophy* (Cambridge, 1938), vol. II, Pt. I, pp. 319-323.

¹⁰ *Philosophical Studies* (New York and London, 1934), pp. 140-142 (in "The Relation of Time and Eternity").

or with becoming—but not with both. A decision to side with logic *may* be tantamount to denying time itself. Whether that is indeed the case this paper has not sought to determine. But if it is, then we shall be faced with a choice between logic on the

one hand and the most certain deliverances of experience on the other. In such a situation, neither logic nor experience may safely be taken at face value.

Received September 3, 1969

Drew University

X. EMPIRICISM AND ETHICAL REASONING

EDWARD F. WALTER

MOST American and English philosophers in the empiricist tradition—those who believe that the scientific method alone attains truth—consider ethical reasoning to be nonfactual (non-scientific). Earlier in this century logical positivists argued that science and ethics were separated on the grounds that ethical judgments were emotive expressions, while scientific statements were verifiable propositions. Later, this view was modified both by emotivists and those who proposed that ethical reasoning employed unique processes that cannot be reduced to emotivism or science. According to the latter group, earlier theoreticians failed to recognize the autonomy of ethics.

In this paper I shall disagree with all of the aforementioned views. I shall focus my attention on the variations of these views developed by David Hume, Charles Stevenson, Paul Edwards, and R. M. Hare. All separate science and ethics. After Hume, each increases the role of reason in ethics. However, I shall argue that each at a crucial step in his argument reverts to an untenable emotivism or begs the question.

The belief that the moral philosopher can only describe ethical language and/or reasoning stands on the assumption that ethical judgments are nonfactual. I use the term "assumption" because most of the philosophers holding this viewpoint do not spend much time proving that which is essential to their position. It is taken as "a given," as something "obvious." I contend that this assumption turns out to be a prejudice.

I

The most famous historical statement of this belief was made by David Hume who complained that the moral philosopher surreptitiously changed his verb form from "is" to "ought," thereby introducing a new and unjustified relation.

Hume did not believe that his rule was justified solely because the verb "ought" was used instead of "is," nor solely because evaluations concern behavior. I am sure that he knew that facts can be

expressed with a variety of verb forms. If certain conditions which he specified were met, he also would have attributed a factual status to the ethical statement. Hume denied the factuality of ethical utterances in four steps:

(1) Facts are derived through reason, which is *inactive* in all its forms.¹ The mind is concerned solely with perceptions, which are divided into impressions and ideas. The idea is a copy of an impression.

(2) In perception, values are not the object of sensation. If one observes an act which is considered evil, e.g., murder, one never perceives the evilness of the act, only the sense impressions constituting the act.²

(3) Since values cannot be derived through the passive process of reason, which is proved by (1) and (2), and since we know that values are concerned with actively directing behavior, they must be the product of an active process.

(4) Since valuing is an active process and emoting is the only active process attending all moral judgments, moral judgments are nothing but the expression of emotions.

Statement (2) amounts to Hume's rejection of platonic forms—all knowledge is obtained through sense experience; platonic forms are not objects of sense experience; therefore, platonic forms are not objects of knowledge. Here, "the good" is treated as a platonic form or a derivative of one. Presumably, if we directly experienced the "good" or the "bad," Hume would have believed in a factual moral judgment.

Hume was unwilling to accept an alternative moral fact because he believed that emotions, which are active, are necessary to evaluations, while facts are derived through passive reason.

Contemporary science accepts the former but creates doubts about the latter. It is more probable that our perception of the world is the result of an active process. Consider the following:

(1) The individual who hears a symphonic piece for the first time hears a jumble of sound. The music

¹ David Hume, *A Treatise on Human Nature* (Great Britain, 1958), pp. 457-458.

² *Ibid.*, p. 468.

has no order. Only training—which involves directing attention—leads to an ordered experience.

(2) A conductor *hears* more of the music than a casual listener. If there are exceptions to this rule, they are rare. I am not suggesting that the conductor merely characterizes what he hears differently, as in the case where experience leads him to admonish a player's late entry; I am suggesting that most listeners do not even discriminate the variety of sounds that the conductor does.

(3) Tests with individuals who saw for the first time after operations for the removal of cataracts indicate that visual acuity is a learned skill. All of those tested "saw" the world as an indiscriminate jumble at first. After considerable training, only a small number made the visual discriminations that most people born with sight make.³

(5) Cultural differences bear strongly on the perception. As communication among nations improves, differences are minimized. Nevertheless, Westerners, as a rule, have difficulty both perceiving and interpreting Eastern art forms.

These few examples underscore the proposition that our perception of the world is actively determined by training, conditioning (which I differentiate from training in that it may not be intentional), and beliefs. At the very least, these examples establish that there is not one way of perceiving the world—facts are not obtained passively as Hume believed. In light of these facts, the retention of the Humian rule can be justified only if more adequate reasons were given in its behalf by later adherents.

II

C. L. Stevenson gained fame in the 1930's and 1940's by presenting a version of the emotive theory which, according to devotees, did justice to Hume's empiricism, yet left room in moralizing for reason. He asserted that reasoning occurs in ethical disputes as long as the disputant's basic moral principles are not brought into question. If basic principles conflict, most probably a resolution of the conflict is impossible, even if both dis-

putants intend to be rational, for basic moral principles are expressions of attitudes.⁴

Unlike earlier emotivists, Stevenson does not make a complete, sharp separation between reason and attitude. ("Attitude" replaces "emotion" as the key term in his work.) He asserts that ethical thinking involves a complicated interweaving of beliefs and attitudes. However, his definition of the term "attitude" is quite unilluminating in view of its importance in his work.

A precise definition of "attitude" is too difficult a matter to be attempted here; hence the term, central though it is to the present work, must for the most part be understood from its current usage, and from the usage of the many terms ("desire," "wish," "disapproval," etc.) which name specific attitudes.⁵

It is important to know whether or not attitudes are direct, unmediated expressions of the biological structure, developed solely through conditioning and training, or a combination of both. If it is the first, then reason is important in altering attitudes. This would mean that reason could only correct factual mistakes about the objects of interest, find means to attain them, etc. If it is the second or the third, since these factors can be altered, then it is theoretically possible for all ethical disputes to be resolved rationally or by appeal to fact. Stevenson, in introductory passages, talks in the second and third way, but when he presents evidence to support his thesis, he reverts to the first claim. An example will illustrate my point.

Some ethical disagreements *seem* rooted . . . in temperamental differences, as when an oversexed, emotionally independent adolescent argues with an undersexed, emotionally dependent one about the desirability of free love. In these cases, the growth of science may, for all that we can now know, leave ethical disagreement permanently unresolved.⁶

This assertion suggests that no amount of knowledge can make someone who is constitutionally undersexed or oversexed react differently to sexual stimuli. This is Humian claim in modern language—at least some basic moral judgments are the

³ J. Z. Young, *Doubts and Certainty in Science: A Biologist's Reflections on the Brain* (New York, 1960).

⁴ Hume did not strictly adhere to an emotive theory of ethics. Several years ago, A. C. MacIntyre attempted to develop a consistent interpretation of Hume. (*Philosophical Review*, vol. 68 [1959]). Most critics have found MacIntyre's argument wanting. Stevenson himself makes passing references to the inconsistencies in Hume. (Pp. 273-276.)

Stevenson did not present the first twentieth-century version of the emotive theory. Others by C. K. Ogden, I. A. Richards, and A. J. Ayer preceded his.

⁵ C. L. Stevenson, *Ethics and Language* (New Haven and London, 1962), p. 60.

⁶ *Ibid.*, pp. 136-137.

direct expression of biological conditions without the mediation of training and knowledge.

The only trouble with Stevenson's argument about free love is that I am unsure who favors free love and who opposes it. Haven't all of us met people like the following?

(1) Individual *A* is a member of the Free Love Association of America and spends most of his time propagandizing for it. The odd thing about *A* is that he seems to be asexual himself.

(2) Individual *B* is a student of *A*'s. After *A* convinced *B* of the desirability of free love, *B* joined the "jet set" and now spends most of his time happily compiling a modern-day Leporello's Catalogue.

(3) Individuals *D* and *C* correspond in physical endowment to *A* and *B*, respectively. Both being members of the Roman Catholic Church, they oppose free love on religious grounds. However, *D* is the "happy-go-lucky" president of a local Knights of Columbus, while *C* is a neurotic, nerve-wrecked frequenter of the confessional, who in his spare time wrote the song, "Somewhere There Must Be a Better Place Than This."

In other words, there seems to be a gap between physical conditions and evaluations which is filled by cultural bias, training, and beliefs. If we want to consider moral judgments to be unmediated expressions of biological conditions, then we are back with Hume's separate processes. If we believe that natural drives are affected by the physical and social environment, then physical differences cannot be a ground for denying that ultimate moral agreement can be attained rationally.

Structural differences in physical organisms guarantee the development of temperamental differences, but this does not logically imply moral differences. It can be argued that a rational morality seeks to control the environment so that people with divergent temperaments can operate in harmony. Harmony may be desirable as a precondition for the attainment of "self-interest."

A second problem for his theory is that he merely assumes that attitudes are non-factual in spite of his promise to demonstrate that reason is a limited tool in resolving ethical disputes.⁷

He begs the question, e.g., in his discussion of a personal decision to change one's own values. The example he uses is of an individual who abandons a

value which he accepted as a God-given rule after he becomes an agnostic. He argues that it cannot be concluded that the factual change caused the value change, for the normative decision to change the value upon doubting the fact requires the application of a logically prior attitude.⁸

But this assumes that at least one attitude is independent of beliefs and that the role of rationality is to order attitudes. He offers no reason to support this claim, other than the fact that we can always find latent attitudes. But we also find latent beliefs. Why cannot the latent attitude be based on a latent belief? There is no logical reason to reject the claim that every attitude in the ethical process is not itself based on a conjunction of facts.

A third reason for rejecting a factual resolution of moral disagreements is made by Stevenson in discussing a dispute about the desirability of premarital intercourse. He argues that despite the seeming ease with which the conflict can be resolved, a closer investigation turns up the fact that it is an exceedingly difficult problem which leads most disputants to abandon rational methods of resolution.⁹

But this proves only that a rational resolution of the problem is difficult, not impossible. What he proves is what is already known—that most people desert the rational method in ethical disputes. To substantiate his theory he must prove that it is impossible or highly improbable that ethical disputes can be rationally resolved.

The difficulties which Stevenson suggests stand in the way of a rational resolution of ethical disputes, encumber the resolution of scientific disputes as well.

The scientist who is breaking new ground must separate relevant evidence from irrelevant evidence and distinguish that which he has good reason to assume from that which requires evidence. In resolving these problems, disputes may occur about the verifiability of scientific claims and the role of theory in establishing fact. The ordinary scientist resolves these disputes by appeals to authority, intuition, etc., the very way ethical disputes are resolved, according to Stevenson.

In conclusion, I believe that it is not unfair to say that Stevenson has demonstrated only that most people do not employ rational means of resolving ethical disagreements—which has never been doubted. He did not prove that these problems

⁷ Stevenson, *op. cit.*, p. 31.

⁸ *Ibid.*, p. 132.

⁹ *Ibid.*, pp. 124-125.

could not be resolved rationally, which was his intention.

III

According to some philosophers, Paul Edwards' version of the emotive theory improved upon earlier versions in that he more adequately described the uses of reason in ethics. Edwards criticizes Stevenson for limiting a resolution in ethical disputes to instances where disputants have come to hold the same view for any psychological reason whatever or because either disputant has made a factual mistake which is clarified.¹⁰ Edwards argues that ethical discussions can lead to an investigation of the truth or falsity of the *reasons* offered for the moral judgment, and that the outcome of this part of the discussion would have bearing on the value judgments, whether or not the disputants recognized it.¹¹

Further on, Edwards states that the limits of rationality are reached in disputes in which one or both disputants make "fundamental" moral judgments, i.e., when one or both are "unable or unwilling to support a moral judgment with anything that would be considered a reason."¹² In such cases, the moral judgment has "emotive meaning" only—it serves to express the attitude of the speaker.

The notion of a fundamental moral judgment is crucial to Edwards' work. In his exposition, two important difficulties arise in relation to it: (1) the examples he gives of a fundamental judgment seem to be the same kind that he refuses to consider fundamental in Stevenson's book, and (2) he does not develop criteria for determining when a judgment is fundamental.

In regard to his criticism of Stevenson's claim, he argues that an evaluation implies that certain features are present in that which is evaluated which must be substantiated. The failure of the evaluator to acknowledge this does not stand in the way of the resolution of an ethical conflict. He points out that a dispute about a person's negative evaluation of euthanasia, which is based upon a belief in God's injunction, is settled regardless of the disputant's feelings if agnosticism be successfully defended.¹³

Edwards differs from Stevenson in these passages because he suggests that ethical attitudes develop because of beliefs; they are not, as Stevenson implies, direct unmediated expressions of biological conditions. However, his example of a fundamental moral judgment reverts to Stevenson's emotivism. He recalls a dispute between two women about the morality of stealing in which one abandons supporting her judgment with reason on the grounds that stealing is *just* wrong and that no reason would sway her from this belief.¹⁴

In this case, he asserts that the judgment is fundamental simply because the woman refuses to give reasons or listen to reasons concerning her judgment. How does this differ from the euthanasia case? Does Edwards doubt that judgments like the one made by this woman develop because of reasons? Knowledge of sociology and psychology convince us that such judgments develop in a cultural milieu and as a consequent of specific environmental forces. The fact that the woman fails to acknowledge that reasons are relevant to her judgment does not mean that they are not. Her refusal to give reasons proves her irrationality, but it does not prove that moral judgments must be irrational.

Edwards' conception of a fundamental moral judgment is inadequately explicated in his book. In light of his criticism of Stevenson, a fundamental moral judgment would be one which is necessarily independent of the beliefs of the evaluator. But the criteria he suggests to locate a fundamental judgment—that one is unable or unwilling to give reasons in support of his judgment—is insufficient to achieve this end for the inability and unwillingness might be the result of (1) intellectual crassness, (2) training which leads one to believe that he ought to express his feelings without intellectualizing about them, and/or (3) irrationality. These factors may encumber resolutions of ethical disputes, but they encumber resolutions of scientific disputes as well. And Edwards was trying to demonstrate the differences between scientific and ethical reasoning.

The possibility of a science of ethics would be destroyed if a fundamental moral judgment is the direct expression of an attitude that *could not* be

¹⁰ Paul Edwards, *The Logic of Moral Discourse* (Glencoe, Illinois, 1955), pp. 180–181.

¹¹ *Ibid.*, pp. 36–41. Excellent examples of this type of argument are contained in the book—one such example is presented on pages 172–178. I shall discuss it later.

¹² *Ibid.*, pp. 182–183.

¹³ *Ibid.*, p. 179.

¹⁴ *Ibid.*, pp. 184–185.

changed because it is independent of the individual's knowledge and training. If it is dependent upon knowledge and training, then new information is relevant to its continuance. If it is merely pointed out that an individual *refuses* to approach the attitude rationally, then it can only be asserted that the individual *is* irrational—not that basic moral judgments *are* irrational.

IV

Most philosophers in the empiricist tradition would admit that the emotive theory in any form presented insurmountable problems. Applying Wittgenstein of the *Philosophical Investigations* to the problem, they have argued that the mistake of the logical positivist in ethics was that he demanded that ethical language adhere to certain prescribed expectations, while as a matter of fact ethical language has a logic of its own. It is, in some sense, like scientific language; in some sense, like emotive language; and, in toto, unique. In this section I shall consider the work of one of the most famous philosophers in this movement, R. M. Hare.

Against those who would propose a science of ethics, Hare would argue that science is concerned with description while ethics is concerned with prescription. He defends this claim by pointing out that ethical reasoning, unlike scientific reasoning, always involves an ethical premiss.¹⁵ A perspicacious investigation of any ethical argument will turn up an overt or covert value—one never finds an instance of ethical reasoning that is derivable only from a fact or conjunction of facts. Accepting an ethical judgment requires that one make a "decision in principle" about it.¹⁶

My doubt about his theory stems from the fact that it stands on the admitted fact that one does not ordinarily observe a moral reasoning process which is closed. To find out whether or not fundamental moral principles are factual, one would have to give a "complete specification of a person's way of life" which remains for Hare only a possibility, not a probability.¹⁷ But since his conclusion rests on partial evidence, he begs the question. That a factual derivation has not been demonstrated does not prove that it cannot be.

All that he has shown is that our value systems

are open and immensely complicated. In a debate about values, it is true that some value or values will be pre-supposed. This does not mean that they themselves cannot be justified in another context. As I have pointed out in the discussion of Stevenson, the uncovering of latent assumptions turns up factual considerations as well. It then becomes problematic to determine which are logically prior.

Hare's critiques of other philosophical attempts at bridging the "fact-value" gap demonstrate similar question begging. Consider his rejection of Stephen Toulmin's ethical end: the attainment of peace and harmony.

Suppose that someone were disputing this, by saying. "Without conflict, the full development of manhood is impossible; therefore it is a bad reason for calling a practice right to say that it would involve the least conflict of interests." We might reply, as Mr. Toulmin does here, "This seems so natural and intelligible . . . What better kinds of reason could you want?" And if we said this, and the other man replied, "I don't find it natural or intelligible at all; it seems to me that development of manhood is a cause superior to all others, and provides the only good reason for any moral conclusion," then it would be clear that what was dividing us was a moral difference.¹⁸

In this passage, Hare *assumes* that for some people the dispute, no matter what is said, will leave the dispute unresolvable on factual grounds because after both disputants were aware of all of the facts they might make different "decisions in principle." This may be so, but it cannot be assumed to be so.

It must be clear that what we value is affected by training and beliefs. It is not accidental that graduates of military academies tend to express Hare's demurrer from Toulmin's pacifism. Furthermore, such beliefs arise in contextual situations which promise survival for the strong.

A complete specification of the life of the warrior might turn up the facts that he believes (erroneously) that it is natural for man to be aggressive, when in fact it is only natural for man to exercise; that he believes (erroneously) that only the aggressive state survives, when in fact survival depends upon a variety of conditions including some which are beyond the control of the citizens of the state; he believes (erroneously) that compromises among nations are impossible because of the "machia-

¹⁵ R. M. Hare, *The Language of Morals* (New York, 1964), pp. 175-179; and again in his criticism of Toulmin's theory in his review of Toulmin's *The Place of Reason in Ethics* in *The Philosophical Quarterly*, vol. 1, (1951), p. 374.

¹⁶ Hare, *The Language of Morals*, *op. cit.*, Pt. I, chap. 4.

¹⁷ *Ibid.*, pp. 68-69.

¹⁸ Hare, review of Stephen Toulmin's *The Place of Reason in Ethics*, *op. cit.*, p. 374.

velian" nature of man, when in fact this only appears so because men, as a rule, do not attempt to resolve their problems honestly.

At this juncture, I suppose that a follower of Hare would want to interject the claim that if the soldier's factual picture has changed, he must make a decision in principle about how this new knowledge will affect him, which involves employing a moral principle. This rejoinder misses my point. I am claiming that if a moral judgment is upset, the employment of a more basic moral principle *may* imply *latent* facts about the more basic moral principle.

Hare has not shown that moral decisions in principle are not based on factual considerations. The fact that we always find a moral principle in a moral argument does not rule out that behind the latent moral principle, latent facts, without which the principles would be absent, might be found.

V

All of the philosophers under discussion proposed the separation of science and ethics on the grounds of the limits of rationality in ethical matters. Hume is the only one who completely separates ethical activity from rational activity. Stevenson, Edwards, and Hare are willing to grant that rationality has some bearing on ethical processes, but they do not specify clearly the limitations of rationality. When they talk of "fundamental moral principles," or "basic attitudes," they either resort to an emotivism—these values are natural to man and cannot be affected by rational processes—or to the suggestion that individuals refuse to deal rationally with them which does not distinguish science from ethics.

When they talk about "fundamental moral values" or "temperamental differences," they suggest biological differences among people which cannot be modified, changed, or directed through

rational means. If they do not imply this, if they imply that biological differences undergo cultural development, then rationality has not been shown to be impotent in these matters. Cultural development is a learning process, which can be modified or changed with reason.

However, I suspect that these philosophers imply the former. I suspect that they imagine that some natural biological drives, which vary from man to man, are directly expressed as value judgments. After Freud, it is common to talk of biological or natural drives, such as the sex drive, the aggressive drive, the death wish, etc. I doubt if this kind of talk is justified—unless one resurrects "innate ideas." An example might make my objection clear.

In contemporary literature, the "aggressive drive" is used by many to prove that man will ultimately destroy himself in a hydrogen war. As far as I have read, I have never seen a good argument presented to substantiate this view. It is true that man's biological structure is such that he undergoes internal stresses that tend to be expressed externally, but there is no necessary external act that relieves the internal stress. Some join the army; others become professional athletes; still others run around the block. The way one deals with basic drives depends upon the opportunities open to him for expression, the beliefs he has about them, etc. Since the avenues of expression are innumerable, it is not unreasonable to believe that we can determine through rational processes how to express our internal states. If the usual means of self-expression are culturally determined, they can be altered in the light of new information.

In short, the suggestion that certain *values* are "natural" to man has never been proven. What we do have evidence for is the claim that natural drives are expressed in a variety of ways and when coupled with beliefs form the basis of values. But this calls for a rational direction of natural urges through an understanding of self and the social and physical environment.

DISSERTATION ESSAY COMPETITION

The Publisher of *The Review of Metaphysics*, The Philosophy Education Society, Inc., announces an Essay Competition. The Competition is open to participants who have been awarded the Ph.D. degree in Philosophy in the United States or Canada during 1970. Essays must be a Chapter from a Dissertation, or a Paper based directly upon the Dissertation. Essays on any topic dealt with in the Dissertation are acceptable. The author of the prize-winning Essay will receive \$100 and it is expected that the Essay will be published in *The Review of Metaphysics* shortly after the Award. The best Essay will be selected by the following Committee: Edwin B. Allaire, University of Texas; Robert Fogelin, Yale University; Charles H. Kahn, University of Pennsylvania, and the Editor of *The Review of Metaphysics*, Richard J. Bernstein, Haverford College. Essays may be sent any time before January 15, 1971 to *The Review of Metaphysics*, Lyman Beecher Hall, Haverford College, Haverford, Pennsylvania 19041 U.S.A. Essays should be marked as submissions for the Competition. The winning Essay will be announced by April 1, 1971. Additional inquiries concerning the Competition may be directed to the above address.

BOOKS RECEIVED

- AYER, A. J., *Metaphysics and Common Sense* (San Francisco: Freeman, Cooper & Company, 1970), pp. 267.
- BAHM, ARCHIE J., *Directory of American Philosophers* (Albuquerque, New Mexico: University of New Mexico, 1970), pp. 436. \$14.95.
- BRAND, MYLES (ed.), *The Nature of Human Action* (Illinois: Scott, Foresman & Company, 1970), pp. 342.
- BRODY, BARUCH A. (ed), *Moral Rules and Particular Circumstances* (Englewood Cliffs, New Jersey: Prentice-Hall, 1970), pp. 181.
- CASTELLI, ENRICO (Direttore), *Bibliografia Filosofica Italiana* (Rome: Edizioni Abete, 1969), pp. 644.
- CASTELLI, ENRICO (Direttore), *Il Senso Comune* (Italy: Padova, 1970), pp. 186.
- CAHN, STEVEN M., *The Philosophical Foundations of Education* (New York: Harper & Row, Publishers, 1970), pp. 433.
- DILWORTH, DAVID A. and HIRANO, UMEYO, *An Encouragement of Learning* (Tokyo: Sophia University, 1969), pp. 128. \$5.75.
- LIDIN, OLOF G., *Distinguishing the Way* (Tokyo: Sophia University, 1970), pp. 139. \$7.25.
- MORICK, HAROLD, *Introduction to the Philosophy of Mind: Readings from Descartes to Strawson* (Illinois: Scott, Foresman & Company, 1970), pp. 315.
- PAHEL, KENNETH and SCHILLER, MARVIN (eds.), *Readings in Contemporary Ethical Theory* (Englewood Cliffs, N. J.: Prentice-Hall, 1970), pp. 572. \$8.95.
- PARWEZ, G. A., *Islam: A Challenge to Religion* (Lahore, Pakistan: Zarreen Art Press, 1968), pp. 392.
- Studia Leibnitiana, Suppl. IV—Theologie-Ethik* (Wiesbaden, Germany: Franz Steiner Verlag GMBH, 1969), pp. 263.
- TAVANEC, P. V. (ed.), *Problems of the Logic of Scientific Knowledge* (Dordrecht, Holland: D. Reidel Publishing Company, 1970), pp. 429.
- THEAU, JEAN, *La Conscience de la Duree et le Concept de Temps* (Toulouse, France: Edouard Privat, editeur, 1969), pp. 311.
- THEAU, JEAN, *La Critique Bergsonienne du Concept* (Toulouse, France: Presses Universitaires De France, 1968), pp. 620.

CORRIGENDA

Volume 6 (1969)

HECTOR-NERI CATASTANEDA: "*Ought, Value, and Utilitarianism*," (pp. 257-275).

Page 257, col. 2, line 2 from bottom: Read ' $(cO)_3$ ' for ' cO_3 '

Page 261, col. 1, lines 10-15 from bottom: Delete square brackets

Page 265, col. 2, line 32: read 'ge' for 'e'

Page 266, col. 1, line 19: Read 'I' for 'I'

line 4 in Case No. I: Read ' $\sim A$ ' for the first occurrence of 'A'

line 5 in Case No. I: Read ' $\sim B$ ' for the first occurrence of 'B'

Page 268, col. 1, line 32: Read 'I' for 'I'

line 41: Delete the first occurrence of the ' \sim '

col. 2, line 21: Read ' (o_2) ' for ' (o_2) '

lines 23-24: Delete ' $V(\sim A \ \& \ B \vee \sim A \ \& \ \sim B) \sim \phi(x,y); V(A \ \& \ \sim B \vee$
 $\sim A \ \& \ B) = \phi(y,z);$ '

line 26: Read ' $\phi(x,w)$ ' for ' $\phi(w,x)$ '

Page 270, col. 2, line 3 from bottom: Read Jeffrey for Jeffrey's

W. GREGORY LYCAN: "*On Intentionality and the Psychological*" (pp. 305-311).

Page 308, col. 2, line 20: The sentence should read: since ' S knows that q & not- q ' entails.

AMERICAN PHILOSOPHICAL QUARTERLY

MONOGRAPH SERIES

Edited by NICHOLAS RESCHER

The *American Philosophical Quarterly* also publishes a supplementary Monograph Series. Apart from the publication of original scholarly monographs, this is to include occasional collections of original articles on a common theme. The current monograph is included in the subscription, and past monographs can be purchased separately but are available to individual subscribers at *half price* (though not to institutional subscribers).

- No. 1. STUDIES IN MORAL PHILOSOPHY. *Contents*: Kai Nielsen, "On Moral Truth"; Jesse Kalin, "On Ethical Egoism"; G. P. Henderson, "Moral Nihilism"; Michael Stocker, "Supererogation and Duties"; Lawrence Haworth, "Utility and Rights"; David Braybrooke, "Let Needs Diminish That Preferences May Prosper"; and Jerome B. Schneewind, "Whewell's Ethics." 1968, \$6.00.
- No. 2. STUDIES IN LOGICAL THEORY. *Contents*: Montgomery Furth, "Two Types of Denotation"; Jaakko Hintikka, "Language-Games for Quantifiers"; James W. Cornman, "Types, Categories, and Nonsense"; Robert C. Stalnaker, "A Theory of Conditionals"; Alan Hausman and Charles Echelbarger, "Goodman's Nominalism"; Ted Honderich, "Truth: Austin, Strawson, Warnock"; and Colwyn Williamson, "Propositions and Abstract Propositions." 1968, \$6.00.
- No. 3. STUDIES IN THE PHILOSOPHY OF SCIENCE. *Contents*: Peter Achinstein, "Explanation"; Keith Lehrer, "Theoretical Terms and Inductive Inference"; Lawrence Sklar, "The Conventionality of Geometry"; Mario Bunge, "What Are Physical Theories?"; B. R. Grunstra, "The Plausibility of the Entrenchment Concept"; Simon Blackburn, "Goodman's Paradox"; Stephen Spielman, "Assuming, Ascertaining, and Inductive Probability"; Joseph Agassi, "Popper on Learning from Experience"; D. H. Mellor, "Physics and Furniture"; and Michael Slote, "Religion, Science, and the Extraordinary." 1969, \$6.00.
- No. 4. STUDIES IN THE THEORY OF KNOWLEDGE. *Contents*: John Knox, Jr., "Do Appearances Exist?"; Norman Malcolm, "Wittgenstein on the Nature of Mind"; W. Donald Oliver, "A Sober Look at Solipsism"; John L. Pollock, "The Structure of Epistemic Justification"; Frederick Stoutland, "The Logical Connection Argument"; Peter Unger, "Our Knowledge of the Material World"; Alan R. White, "What Might Have Been." 1970, \$6.00.

THE UNDERGRADUATE JOURNAL OF PHILOSOPHY

Founded and edited by students at Oberlin College, the *Journal* provides a forum for the presentation and criticism of philosophical articles, discussions, and book reviews written by undergraduates. Not committed to any branch or school of philosophy, the *Journal* will continue to publish works on a wide variety of philosophically challenging subjects

MANUSCRIPTS should be typed and not exceed 4000 words.

SUBSCRIPTIONS:

STUDENTS: \$1.00 INDIVIDUALS: \$1.50
INSTITUTIONS: \$2.00

Published in December and May.

Send all correspondence to:
King Building 105, Oberlin, Ohio 44074

British Journal for the Philosophy of Science

Volume 21, Part 3, August 1970

DINGLE Causality and Statistics in Modern Physics

GOROVITZ Inscriptionalism and the Objects of Explanation

GIEDYMIN The Paradox of Meaning Variance

CLEAVE The Notion of Validity in Logical Systems with Inexact Predicates

20s. net (U.S.A. \$3.00). Annual subscription
60s net (U.S.A. \$9.50 for four issues.

CAMBRIDGE UNIVERSITY PRESS
Bentley House, 200 Euston Road,
London, N.W.1.
American Branch: 32 East 57th Street,
New York, N.Y. 10022.

DIALOGUE

*Canadian Philosophical Review—Revue Canadienne
de Philosophie*

Editors: VENANT CAUCHY and MARTYN ESTALL

VOL. IX—1970—No. 2

ARTICLES

Hegel: Time and Eternity	KLAUS HEDWIG
Time in Hegel's Philosophy	M. E. WILLIAMS
Travail et téléologie chez Hegel selon Lukács	YVON BLANCHARD
La Pensée politique de Comte et de Hegel	OLIVIER REBOUL
Le Rapport entre subjectivité et société civile	GABRIEL KORTIAN

NOTES	DISCUSSIONS	REVIEWS
-------	-------------	---------

Subscriptions: \$10.00 a year to individuals; \$12.00 to libraries.
Payable to the Canadian Philosophical Association in care of Norman J. Brown, Department of Philosophy, Queen's University, Kingston, Ontario.

Progress and Regress in Philosophy

Volume I

LEONARD NELSON
Translated by N. H. PLAMER

This book consists of a translation of the Introductory Remarks and Part One (on Hume and Kant) of Nelson's *Fortschritte und Rückschritte der Philosophie*. Volume II of this work is in preparation and will be announced in due course.

60s. (£3.00) net

Categorical Frameworks

STEPHAN KORNER

Philosophers, anthropologists and historians of ideas have often noted a close connection between men's classificatory schemes, their standards of intelligibility and their metaphysical convictions. Professor Korner examines this connection in some detail from a logico-philosophical point of view.

Cloth 45s. (£2.25) net
Paper 25s. (£1.25) net

BASIL BLACKWELL

PRINTED IN ENGLAND

by C. Tinling & Co. Ltd., Liverpool, London and Prescott